

**Thermodynamics of the dimerization  
of amyloidogenic proteins:  
A computer simulation study using  
Stochastic Approximation Monte Carlo**

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften  
(Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät II  
Chemie, Physik und Mathematik

der Martin-Luther-Universität  
Halle-Wittenberg

vorgelegt von

Herrn Christian Sommerfeld

geb. am [REDACTED] in [REDACTED]



**Gutachter:**

Prof. Dr. rer. nat. Wolfgang Paul

Prof. Dr. rer. nat. Dieter Sebastiani

Prof. Dr. rer. nat. Michael Bachmann

**Tag der öffentlichen Verteidigung:**

Datum: 7. Juli 2025

Uhrzeit: 14:00 Uhr

Ort: Seminarraum 1.23,  
von-Seckendorff-Platz 1,  
06120 Halle (Saale)



For Karl.



# Abstract

This thesis explores the dimerization of amyloidogenic proteins – polyglutamine, amyloid-beta ( $A\beta$ ), and parathyroid hormone (PTH) – using computational methods to uncover key thermodynamic and structural properties. Simulations characterize phase transitions and molecular structures within distinct phases. For polyglutamine, the dependence of phase transition temperatures on chain length and aggregate formation is analyzed.  $A\beta(1-40)$  and  $A\beta(1-42)$  systems reveal structural and thermodynamic differences driving aggregation. PTH of various length is studied to assess chain length effects on dimerization and aggregation. A comparison between  $A\beta$  and PTH dimerization is conducted. The intermediate-resolution PRIME20 protein model and stochastic approximation Monte Carlo (SAMC) simulations are employed. This work advances understanding of the thermodynamic and structural principles underlying amyloid dimerization and aggregation.

---

# Zusammenfassung

Diese Arbeit untersucht die Dimerisierung amyloidogener Proteine – Polyglutamin, Amyloid-Beta ( $A\beta$ ) und Parathormon (PTH) – mittels computergestützter Methoden, um zentrale thermodynamische und strukturelle Eigenschaften aufzudecken. Simulationen charakterisieren Phasenübergänge und molekulare Strukturen in verschiedenen Phasen. Für Polyglutamin wird der Zusammenhang zwischen Phasenübergangstemperaturen und Kettenlänge sowie der Aggregate untersucht.  $A\beta(1-40)$ - und  $A\beta(1-42)$ -Systeme zeigen strukturelle und thermodynamische Unterschiede, die die Aggregation bestimmen. PTH mit verschiedenen Kettenlängen wird analysiert, um deren Einfluss auf Dimerisierung und Aggregation zu bewerten. Ein Vergleich zwischen  $A\beta$  und PTH Dimerisierung wird durchgeführt. Das PRIME20-Proteinmodell, welches intermediäre Auflösung hat, sowie stochastische Monte-Carlo-Simulationen (SAMC) werden eingesetzt. Diese Arbeit liefert neue Einblicke in die Prinzipien der Amyloid-Dimerisierung und -Aggregation.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Research questions and aims . . . . .	4
1.2	Structure of the thesis . . . . .	5
<b>2</b>	<b>Protein Model</b>	<b>6</b>
2.1	Proteins . . . . .	6
2.1.1	Amyloid fibrils . . . . .	9
2.2	Modeling of proteins . . . . .	9
2.3	PRIME20 protein model . . . . .	11
2.3.1	Geometry and interaction potentials . . . . .	11
2.3.2	Parameters . . . . .	16
<b>3</b>	<b>Simulation Method</b>	<b>20</b>
3.1	Statistical mechanics . . . . .	20
3.1.1	Statistical ensembles . . . . .	20
3.2	Standard Monte Carlo methods . . . . .	22
3.2.1	Simple sampling Monte Carlo . . . . .	22
3.2.2	Importance sampling Monte Carlo . . . . .	23
3.3	SAMC . . . . .	25
3.3.1	Algorithm of SAMC . . . . .	26
3.3.2	Implementation details and simulation parameters . . . . .	27
3.3.3	Monte Carlo moves . . . . .	31
3.4	Observables . . . . .	32
3.4.1	Transition between ensembles . . . . .	32
3.4.2	Thermodynamic observables . . . . .	33
3.4.3	Structural observables . . . . .	35
<b>4</b>	<b>Dimerization of Polyglutamine</b>	<b>38</b>
4.1	Thermodynamic analysis . . . . .	38
4.2	Structural analysis . . . . .	43
4.2.1	Tensor of gyration . . . . .	43
4.2.2	Hydrogen bond contact probability maps . . . . .	45
4.2.3	Intra- vs. inter-molecular contacts . . . . .	48
4.3	Temperature mapping . . . . .	53
4.4	Conclusions . . . . .	54
<b>5</b>	<b>Dimerization of Amyloid Beta</b>	<b>55</b>
5.1	Thermodynamic analysis . . . . .	55
5.1.1	Temperature mapping . . . . .	58
5.2	Structural analysis . . . . .	60
5.2.1	Tensor of gyration . . . . .	60
5.2.2	Intra vs. inter-molecular contacts . . . . .	61

5.2.3	Hydrogen bond contact probability maps . . . . .	66
5.2.4	Side chain contact probability maps . . . . .	77
5.2.5	Ramachandran plot analysis . . . . .	80
5.3	Conclusions . . . . .	85
<b>6</b>	<b>Dimerization of hPTH</b>	<b>87</b>
6.1	Thermodynamic analysis . . . . .	87
6.1.1	Temperature mapping . . . . .	90
6.2	Structural analysis . . . . .	91
6.2.1	Relative shape anisotropy . . . . .	91
6.2.2	Intra vs. inter-molecular contacts . . . . .	92
6.2.3	Hydrogen bond contact probability maps . . . . .	93
6.3	Conclusions . . . . .	101
<b>7</b>	<b>Summary and Conclusions</b>	<b>103</b>
7.1	Summary of results . . . . .	103
7.2	Comparison of Amyloid Beta and Parathyroid Hormone . . . . .	105
<b>8</b>	<b>Outlook</b>	<b>107</b>
<b>A</b>	<b>Parameters of PRIME20</b>	<b>123</b>
<b>B</b>	<b>Additional Figures for the Polyglutamine Systems</b>	<b>128</b>
<b>C</b>	<b>Additional Figures for the Amyloid Beta Systems</b>	<b>133</b>
<b>D</b>	<b>Additional Figures for the Parathyroid Hormone Systems</b>	<b>140</b>

# List of Figures

2.1	(a) Structure of an $\alpha$ -amino acid and a peptide chain. The amino acid consists of an amino group ( $\text{NH}_2$ ), a carboxyl group ( $\text{COOH}$ ), and an $\alpha$ -Carbon. The side chain (R) is attached to the $\alpha$ -Carbon. (b) The peptide chain is formed by linking amino acids together by peptide bonds. . . . .	7
2.2	Ramachandran plot for L-Ala residues showing how different combinations of $\phi$ and $\psi$ angles correspond to different secondary structure motifs. The Ramachandran plots for other amino acids vary from that of Ala shown here. The blue areas designate combinations of $\phi$ and $\psi$ angles that are legally accessible for the protein backbone, based on calculations using known van-der-Waals radii and bond angles. Dark blue areas reflect conformations without steric overlap; medium blue areas belong to conformations only allowed at the extreme limits of unfavorable atomic contacts; and light blue areas are conformations that are only allowed if some flexibility is permitted in the bond angles. The white areas designate conformations that are not allowed due to steric hindrance. The figure was adapted from Nelson et al. [62]. . . . .	8
2.3	Schematic representation of the cross- $\beta$ -sheet motif in amyloid fibrils. The $\beta$ -strands are oriented perpendicular to the fibril axis. The $\beta$ -strands are connected by HBs between the backbone atoms. The two stacked $\beta$ -sheets are held together by hydrophobic interactions between the side chains in a steric zipper. . . . .	9
2.4	(a) Schematic representation of an $\alpha$ -amino acid residue in the PRIME20 model. The backbone of the amino acid residues is represented by three beads: the green bead represents the amino group (NH), the red bead the $\alpha$ -Carbon ( $\text{C}_\alpha$ ), and the cyan bead the carboxyl group (CO). The side chain is represented by a single gray bead. In (b) two amino acid residues are connected by a peptide bond. Physical bonds between the beads are depicted as white sticks, while pseudo-bonds are shown in black and yellow. . . . .	11
2.5	Schematic representation of the interaction potentials in the PRIME20 model. (a) Bond potential. (b) Hard-sphere potential. (c) Square-well potential. . . . .	12
2.6	Schematic representation of the construction of the virtual positions of the amide hydrogen and the carboxyl oxygen, used for hydrogen bonding. . . . .	13
2.7	Schematic representation of a peptide conformation in the PRIME20 model, where a single CO (cyan) bead has two possible NH (green) beads to form a hydrogen bond (yellow sticks) with. Side chain beads (R) are hidden for visual clarity. . . . .	14
2.8	Classification of amino acids in the PRIME20 model. The Amino acids are represented by their structure formula, their 3- and their 1-letter name. They are grouped based on their side chain properties. Groups and atoms that are able to form side chain hydrogen bonds are marked by purple circles. The figure is adapted from the PhD thesis of Böker (2019) [61]. . . . .	15

3.1	(a) Example of a 2D simulation box divided into 25 cells. (b) The linked List in table 3.1 is used to find the particles in cells 7 and 8. . . . .	29
3.2	Comparison of thermodynamic observables in the configurational (black) and full-energy (blue) microcanonical ensemble and their respective canonical ensembles. (a) Temperature $T$ as a function of energy $E$ . (b) Squared radius of gyration $R_g^2$ as a function of energy $E$ . (c) Canonical heat capacity $C_V$ as a function of temperature $T$ . (d) $R_g^2$ as a function of temperature $T$ . Values of $U$ , $E$ , $T$ and $C_V$ are given in PRIME20 model units, while $R_g^2$ is given in $\text{\AA}^2$ . . . . .	33
3.3	Schematic example of an entropy curve $S(E)$ featuring a convex intruder. $S(E)$ is colored in blue, while the tangent to the convex intruder is colored in red. . . .	35
4.1	Configurational and total-energy microcanonical entropy of the 9 polyglutamine systems of varying chain length. The entropies are shown in (a) as a function of the potential energy $U$ and in (b) as a function of the total energy $E$ . The different systems are shifted along the y-axis for better visibility as they are only determined by the SAMC algorithm up to an additive constant. . . . .	39
4.2	Plots relevant to the thermodynamic analysis of the C2Q16 system. (a) Temperature $T(U)$ derived from the configurational entropy $S(U)$ , (b) temperature $T(E)$ derived from the full-energy entropy $S(E)$ , (c) canonical heat capacity $c_V(T)$ and (d) microcanonical heat capacity $c(E)$ . In (b) and (d) the gray dashed lines mark the energies of the phase transitions. . . . .	40
4.3	Transition temperatures of the polyglutamine systems, determined from canonical and microcanonical analysis, as a function of chain length. On the left (a) results comparing the different methods are shown. On the right (b) the canonical transition temperatures are plotted against $N^{-1/3}$ to show the scaling behavior. The dashed line is a linear extrapolation to infinite chain length. The origin of the temperature scale in Kelvin on the right y-axis of (b) is discussed in section 4.3. The figures are adapted from Lauer et al. [56]. . . . .	42
4.4	Quantities derived from the tensor of gyration as a function of temperature. (a) Eigenvalues $\lambda_i$ of the tensor of gyration $\mathbf{T}_g$ for all PolyQ systems. (b) Radius of gyration $R_g$ , relative shape anisotropy $\kappa^2$ and asphericity $\Delta$ for all PolyQ systems. Units: $[\lambda_i] = \text{\AA}$ , $[R_g] = \text{\AA}^2$ , $[\kappa^2] = 1$ , $[b] = 1$ . . . . .	43
4.5	Hydrogen bond contact matrices for chain length $N = 14$ at four different temperatures. Each cell $(i, j)$ corresponds to a possible contact between the NH bead of residue $i$ and the CO bead of residue $j$ . The color indicates the probability of a hydrogen bond forming between the beads, according to the color scale on the right of the maps. The colors are scaled to the highest value in the map. Therefore, equal colors in different maps do not correspond to the same contact probability. . . . .	46
4.6	HB matrices of C2Q16 at temperatures below the transition temperature, corresponding to the two ordered states. . . . .	47
4.7	HB matrices of C2Q26 and C2Q36 at temperatures below the transition temperature, corresponding to the system's ordered states. . . . .	48
4.8	Configuration snapshots of folded and aggregated states for various systems. Side chain beads are hidden to improve visibility. Bead colors correspond to different backbone bead types, consistent with fig. 2.4(a). The PolyQ systems of the shown structures are: C2Q14 for (a), (b) and (c); C2Q16 for (d); C2Q26 for (e) and (f); C2Q36 for (g) and (h). The figure were taken from Lauer et al. [56]. . . . .	49
4.9	Number of intra- and inter-molecular HBs vs. the temperature, normalized by the number of residues $N$ in the systems. . . . .	50

4.10	The top plot shows the difference between the number of intra- and inter-molecular HBs as a function of temperature. The bottom plot shows the derivative of the top plot. The circles mark the transition temperatures $T^*$ of the systems. . . . .	51
4.11	Number of intra- and inter-molecular side chain contacts vs. the temperature, normalized by the number of residues $N$ in the systems. . . . .	52
4.12	The top plot shows the difference between the number of intra- and inter-molecular side chain contacts as a function of temperature. The bottom plot shows the derivative of the top plot. The circles mark the transition temperatures $T^*$ of the systems. . . . .	53
5.1	Configurational and total-energy microcanonical entropy of both $A\beta$ systems. (a) Entropy $S(U)$ as a function of energy $U$ . Inset: Microcanonical temperature $T(U)$ . (b) Entropy $S(E)$ as a function of total energy $E$ . . . . .	56
5.2	Plots relevant to the thermodynamic analysis of the $A\beta$ systems. (a) Microcanonical temperature $T(E)$ and heat capacity $c(E)$ . (b) Canonical heat capacity $C_V(T)$ . . . . .	56
5.3	Hydrodynamic radius $R_h$ of the $A\beta$ systems as a function of temperature. The circles indicate the values of $R_h$ at the PT temperatures. . . . .	59
5.4	Quantities derived from the tensor of gyration as a function of temperature for both $A\beta$ systems. (a) Eigenvalues $\lambda_i$ of the tensor of gyration $\mathbf{T}_g$ . (b) Radius of gyration $R_g$ , relative shape anisotropy $\kappa^2$ and asphericity $\Delta$ . Units: $[\lambda_i] = \text{\AA}$ , $[R_g] = \text{\AA}^2$ , $[\kappa^2] = 1$ , $[b] = 1$ . Units: $[\lambda_i] = \text{\AA}$ , $[R_g] = \text{\AA}^2$ , $[\kappa^2] = 1$ , $[b] = 1$ . . . . .	60
5.5	Number of intra- and inter-molecular HBs vs. the temperature for the $A\beta$ systems and normalized by the number of residues in the respective systems. (a) Intra-molecular HB contacts $N_{\text{HB,intra}}(T)$ . (b) Inter-molecular HB contacts $N_{\text{HB,inter}}(T)$ . The bottom plots show the respective derivatives. . . . .	62
5.6	Difference between intra and inter-molecular hydrogen bond contacts as a function of temperature. The top plot shows the difference $\Delta N_{\text{HB}}(T) = N_{\text{HB,intra}}(T) - N_{\text{HB,inter}}(T)$ , while the bottom plot shows the derivative of $\Delta N_{\text{HB}}(T)$ . Circles mark the transition temperatures $T^*$ of the Amyloid beta ( $A\beta$ ) systems. . . . .	63
5.7	Intra- and inter-molecular SC energies vs. the temperature for the $A\beta$ systems. (a) Intra-molecular HB contacts $N_{\text{HB,intra}}(T)$ . (b) Inter-molecular HB contacts $N_{\text{HB,inter}}(T)$ . The bottom plots show the respective derivatives. . . . .	65
5.8	Difference between intra and inter-molecular side chain energies as a function of temperature. The top plot shows the difference $\Delta E_{\text{SC}}(T) = E_{\text{SC,intra}}(T) - E_{\text{SC,inter}}(T)$ , while the bottom plot shows the derivative of $\Delta E_{\text{SC}}(T)$ . Circles mark the transition temperatures $T^*$ of the $A\beta$ systems. . . . .	66
5.9	Intra-molecular hydrogen bond contact probability maps for $A\beta(1-40)$ and $A\beta(1-42)$ . The probability scale is logarithmic. The left column shows the $A\beta(1-40)$ system, while the right column shows the $A\beta(1-42)$ system. The first row displays the HBmaps at $T = 0.164$ , the second row at $T = 0.135$ , and the third row at $T = 0.085$ . . . . .	68
5.10	Inter-molecular hydrogen bond contact probability maps for $A\beta(1-40)$ and $A\beta(1-42)$ . The probability scale is logarithmic. The left column shows the $A\beta(1-40)$ system, while the right column shows the $A\beta(1-42)$ system. The first row displays the HBmaps at $T = 0.164$ , the second row at $T = 0.135$ , and the third row at $T = 0.085$ . . . . .	69

5.11	Configuration snapshots from simulations of the $A\beta(1-40)$ system. The beads represent the backbone beads of the residues in PRIME20: NH, $C_\alpha$ and CO. Side chain beads (R) are omitted for better visibility. The beads along the backbone of each of the two chains are colored in a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state. . . . .	74
5.12	Configuration snapshots from simulations of the $A\beta(1-42)$ system. The beads represent the backbone beads of the residues in PRIME20: NH, $C_\alpha$ and CO. Side chain beads (R) are omitted for better visibility. The beads along the backbone of each of the two chains are colored in a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state. . . . .	75
5.13	Intra- and inter-molecular side chain contact probability maps for $A\beta(1-40)$ and $A\beta(1-42)$ at $T = 0.164$ . The color scale is logarithmic. The left column shows the $A\beta(1-40)$ system, while the right column shows the $A\beta(1-42)$ system. The top row displays the intra-molecular SCmaps, while the bottom row shows the inter-molecular SCmaps. . . . .	78
5.14	Map of energy contributions of contacts between side chains of the residues of $A\beta(1-42)$ system when applying the PRIME20 force field. The color scale indicates the energy contribution of the contact, with blue colors representing attractive interactions and red colors representing repulsive interactions. . . . .	79
5.15	Probability of the $\beta$ -sheet state per residue along the chain for both $A\beta$ systems. The probabilities were calculated from the dihedral angles using the Ramachandran plot analysis. The probability is shown at three different temperatures, $T = 0.195$ , $T = 0.164$ , and $T = 0.085$ . . . . .	81
5.16	Probability of the $\alpha$ -helix state per residue along the chain for both $A\beta$ systems. The probabilities were calculated from the dihedral angles using the Ramachandran plot analysis. The probability is shown at three different temperatures, $T = 0.195$ , $T = 0.164$ , and $T = 0.085$ . . . . .	84
6.1	Configurational and total-energy microcanonical entropy of PTH(1-34), PTH(1-42) and PTH(1-84). (a) Entropy $S(U)$ as a function of potential energy $U$ . Inset: Microcanonical temperature $T(U)$ . (b) Entropy $S(E)$ as a function of total energy $E$ . . . . .	88
6.2	(a) Microcanonical temperature $T(E)$ and heat capacity $c(E)$ . The positions of local maxima in $c(E)$ are marked by dashed lines and transferred to $T(E)$ to find the corresponding PT temperatures. (b) Canonical heat capacity $C_V(T)$ . . . . .	89
6.3	Relative shape anisotropy $\kappa^2$ of the three PTH systems. Circles mark the system's PT temperatures. . . . .	92
6.4	Number of intra- and inter-molecular hydrogen bond contacts vs. the temperature for the three PTH systems and normalized by the number of residues in the respective systems. (a) Intra-molecular HB contacts $N_{HB,intra}(T)$ . (b) Inter-molecular HB contacts $N_{HB,inter}(T)$ . . . . .	93
6.5	Difference between intra and inter-molecular hydrogen bond contacts as a function of temperature. The top plot shows the difference $\Delta\mathcal{N}_{HB}(T) = \mathcal{N}_{HB,intra}(T) - \mathcal{N}_{HB,inter}(T)$ , while the bottom plot shows the derivative of $\Delta\mathcal{N}_{HB}(T)$ . Circles mark the transition temperatures $T^*$ of the PTH systems. . . . .	94
6.6	Intra-molecular hydrogen bond contact probability maps for PTH(1-34), PTH(1-42) and PTH(1-84). The probability color scale is logarithmic. The first row displays the HBmaps for PTH(1-34), the second row for PTH(1-42) and the third row for PTH(1-84). The left column shows HBmaps for $T = 0.100$ and the right column for $T = 0.178$ . . . . .	95

6.7	Inter-molecular hydrogen bond contact probability maps for PTH(1-34), PTH(1-42) and PTH(1-84). The probability color scale is logarithmic. The first row displays the HBmaps for PTH(1-34), the second row for PTH(1-42) and the third row for PTH(1-84). The left column shows HBmaps for $T = 0.100$ and the right column for $T = 0.178$ . . . . .	96
6.8	Configuration snapshots from simulations of the PTH systems: (a)–(d) PTH(1-34), (e)–(h) PTH(1-42), (i)–(l) PTH(1-84). Only the backbone beads are shown: NH, C $_{\alpha}$ and CO. Side chain beads (R) are omitted for better visibility. The beads in each chain are colored using a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state. . . . .	100
7.1	Comparison of the number of intra- and inter-molecular hydrogen bonds in the PTH and A $\beta$ systems, relative to the number of beads in the system. Temperature is given in Kelvin, according to the temperature conversions in sections 5.1.1 and 6.1.1. . . . .	106
B.1	PolyQ temperatures . . . . .	128
B.2	PolyQ heat capacities . . . . .	129
B.3	PolyQ canonical heat capacities . . . . .	129
B.4	PolyQ thermodynamic master plot. . . . .	130
B.5	Intra-molecular hydrogen bond contact probability maps for C2Q18, C2Q20, C2Q22, C2Q24 and C2Q28, at $T = 0.140$ . . . . .	131
B.6	Intra- and inter-molecular HBs . . . . .	132
B.7	Intra- and inter-molecular side chain contacts . . . . .	132
C.1	Temperature and inverse temperature vs configurational energy for A $\beta$ (1-40) and A $\beta$ (1-42). . . . .	133
C.2	Amyloid beta hydrogen bond contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42) with a non-logarithmic probability scale. . . . .	134
C.3	Amyloid beta hydrogen bond contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42) with a logarithmic probability scale. . . . .	135
C.4	Amyloid beta intra-molecular hydrogen bond contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42) in the unfolded state at $T = 0.195$ . . . . .	136
C.5	Amyloid beta <b>intra</b> -molecular side chain contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42). . . . .	137
C.6	Amyloid beta <b>inter</b> -molecular side chain contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42). . . . .	138
C.7	Amyloid beta Ramachandran analysis for A $\beta$ (1-40). . . . .	139
C.8	Amyloid beta Ramachandran analysis for A $\beta$ (1-42). . . . .	139
D.1	PTH tensor of gyration. Eigenvalues and other derived properties. . . . .	140
D.2	PTH intra- and inter-molecular side chain energies. . . . .	141
D.3	Difference between intra and inter-molecular side chain energies. . . . .	141
D.4	PTH intra-molecular and inter-molecular side chain contact probability maps for $T = 0.178$ . . . . .	142

# List of Tables

2.1	List of ideal bond lengths ( $b_{ij}$ ), ideal pseudo-bond lengths ( $b_{ij}^{(\text{pseudo})}$ ), bead diameters ( $d_{ij}^{(\text{HS})}$ ) and bond angles of the backbone of the protein chain in the Prime model. Lengths and diameters are given in Å, angles in rad. . . . .	16
2.2	Distances for auxiliary interactions after hydrogen bond formation between beads $\text{NH}_i$ and $\text{CO}_j$ . All distances are given in Å. . . . .	17
2.3	Bond and pseudo-bond lengths between the selected side chains and their backbone beads. Furthermore, the masses of the side chain ( $\text{Mass}^{(\text{R})}$ ) and backbone beads ( $\text{Mass}^{(\text{BB})}$ ) are listed. All lengths are given in Å and all masses in Dalton (Da). . . . .	17
2.4	Energies (square-well depths) for interactions between the selected amino acids. All energies are given in units of $\epsilon_{\text{HB}}$ . . . . .	17
2.5	Side chain bead diameters and square-well widths for interactions between the selected amino acids. All lengths are given in Å . . . . .	18
2.6	Squeeze parameters <i>sqz1–5</i> for pure backbone interactions and how they affect the effective bead diameters. . . . .	18
2.7	Squeeze parameters <i>sqz6–10</i> for interactions between side chains and backbone. Also, the resulting diameters for relevant side chains (values are in Å). . . . .	19
3.1	Head-of-chain array ("HOCL") and Linked List ("LL"). For cell $i$ starting at $\text{LL}[q_1]$ with $q_1$ specified in $\text{HOCL}[i]$ one can follow the trail through LL where each element specifies the index of the next particle in the cell. . . . .	29
3.2	Simulation box side lengths $L$ for the different polymer systems. $N_{\text{res}}$ is the number of residues in the polymer chain, also referred to as the chain length. . .	30
4.1	Phase transition temperatures $T^*$ for all investigated PolyQ systems as obtained from microcanonical and canonical analysis of the entropy $S(E)$ . First order phase transition temperatures in the microcanonical ensemble $T_{\text{micro}}^*$ are determined using the Gibbs construction and second order phase transition temperatures by finding local maxima with positive value in the microcanonical heat capacity. $T_{\text{can}}^*$ is derived from the maxima in the canonical heat capacity. . . . .	42
5.1	Phase transition temperatures $T^*$ of both $\text{A}\beta$ systems as obtained from the microcanonical and canonical analysis. . . . .	57
6.1	Phase transition temperatures $T^*$ of the three PTH systems, obtained via microcanonical and canonical analysis. . . . .	88
A.1	Side chain bond and pseudo-bond lengths (in Å) and bead masses (relative to the mass of $\text{CH}_3$ ). . . . .	123
A.2	Side chain square-well depth (energies). The grouping is in accordance with the Prime20 classification scheme . . . . .	124
A.3	Side chain hard-sphere bead diameter in Å. . . . .	125

A.4	Side chain square-well diameter in Å. . . . .	126
A.5	Changed diameters for interactions between side chains and backbone (after applying squeeze parameters <i>sqz6-10</i> ) for all amino acids. . . . .	127

# Abbreviations

<b>A<math>\beta</math></b> Amyloid beta . . . . .	XI
<b>AD</b> Alzheimer’s disease . . . . .	1
<b>APP</b> amyloid precursor protein . . . . .	2
<b>CG</b> coarse-grained . . . . .	9
<b>cryo-EM</b> cryo-electron microscopy . . . . .	75
<b>DMD</b> discontinuous molecular dynamics . . . . .	11
<b>DOS</b> density of states . . . . .	38
<b>HB</b> hydrogen bond . . . . .	3
<b>HB map</b> hydrogen bond contact probability map . . . . .	36
<b>HOCL</b> Head-Of-Chain list . . . . .	28
<b>HPC</b> high performance computing . . . . .	30
<b>HD</b> Huntington’s disease . . . . .	1
<b>hPTH</b> human parathyroid hormone . . . . .	3
<b>LL</b> Linked-List . . . . .	28
<b>MC</b> Monte Carlo . . . . .	4
<b>MCMC</b> Markov Chain Monte Carlo . . . . .	24
<b>MD</b> molecular dynamics . . . . .	4
<b>mDOS</b> microcanonical density of states . . . . .	25
<b>NMR</b> nuclear magnetic resonance . . . . .	75
<b>PD</b> Parkinson’s disease . . . . .	1
<b>PBC</b> periodic boundary conditions . . . . .	27
<b>PDB</b> protein data bank . . . . .	15
<b>PolyQ</b> polyglutamine . . . . .	2
<b>PRIME</b> Protein Intermediate Resolution Model . . . . .	11
<b>PRIME20</b> Protein Intermediate Resolution Model 20 . . . . .	3
<b>PT</b> phase transition . . . . .	34
<b>PTH</b> parathyroid hormone . . . . .	3
<b>PTHr</b> parathyroid hormone receptor . . . . .	87
<b>REMD</b> replica exchange molecular dynamics . . . . .	58
<b>SAMC</b> Stochastic Approximation Monte Carlo . . . . .	3
<b>SANS</b> small-angle neutron scattering . . . . .	91
<b>SC</b> side chain . . . . .	1
<b>SC map</b> side chain contact probability map . . . . .	77
<b>ssNMR</b> solid-state nuclear magnetic resonance . . . . .	82
<b>WLMC</b> Wang-Landau Monte Carlo . . . . .	25

**Table 1:** 20 proteinogenic  $\alpha$ -amino acids and their respective 3-letter and 1-letter codes.

<b>Amino Acid</b>	<b>3-letter code</b>	<b>1-letter code</b>
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V



# Chapter 1

## Introduction

Death is an inevitable reality that all living beings must face. However, over the past centuries, remarkable advancements in medicine and public health have significantly extended the average human life expectancy, transforming how we perceive aging and health. This increased longevity, while a testament to scientific progress, has been accompanied by a rise in age-related diseases, particularly neurodegenerative disorders. These diseases, marked by a gradual loss of neuronal structure and function, lead to debilitating symptoms such as memory impairment, cognitive decline, and motor dysfunction, profoundly impacting individuals and society.

Among neurodegenerative diseases, Alzheimer’s disease (AD), Parkinson’s disease (PD) and Huntington’s disease (HD), have received significant attention due to their prevalence and devastating effects. AD, in particular, has left a lasting mark on countless lives, either directly or through its impact on loved ones. Despite decades of research and some very recent success in the approval of antibody Alzheimer-therapy<sup>(1)</sup> the underlying molecular mechanisms of these diseases are not yet fully understood, posing significant challenges to the development of effective pharmaceutical strategies. A defining feature of many neurodegenerative diseases is the formation of amyloid fibrils. These are filamentous protein aggregates strongly linked to disease pathology and progression.

The term ”amyloid” was first coined by Rudolf Virchow in 1854 to describe the starch-like appearance of a macroscopic tissue abnormality of cerebral corpora amylacea [2]. Following this misidentification as polysaccharides, Friedrich and Kekulé in 1859 correctly identified the substance as proteinaceous in nature [3, 4]. At the beginning of the 19<sup>th</sup> century, Alois Alzheimer first described amyloid plaques in the brain of a patient with dementia, a hallmark of what would later be known as AD [5]. This added AD to the growing list of diseases associated with the formation of amyloid fibrils, called amyloidoses.

Building on the discoveries of Alzheimer and co., William Astbury pioneered X-ray diffraction studies on amyloid fibrils in the 1930s. His work provided a groundbreaking glimpse into the structural organization of amyloid fibrils, revealing their distinctive stacked beta-sheet architecture [6, 7]. These insights laid the foundation for future structural investigations. Notably, in the 1960s, researchers at the University of Leeds, lead by Geddes et al., established the cross- $\beta$  structure as a definitive structural fingerprint of amyloid fibrils [8]. Their work not only cemented the association of amyloid fibrils with disease but also highlighted their potential physiological roles, reflecting the dual nature of these intriguing protein assemblies.

The cross- $\beta$  structure is the central structural motif of the amyloid fibril. It consists of a series of stacked  $\beta$ -strands arranged perpendicular to the fibril axis, forming a stacked ladder-like structure. The  $\beta$ -strands are separated by a distance of approximately 4.7–4.8 Å, a spacing that arises from the hydrogen bonding between them. Typically, two of these  $\beta$ -sheet ladders are paired and stabilized by a network of intermolecular side chain (SC) interactions. This

---

<sup>(1)</sup>On November 14<sup>th</sup> 2024 the European Medicines Agency (EMA) recommended the granting of marketing authorization for the antibody Lecanemab for the treatment of Alzheimer’s disease [1].

arrangement forms what is known as a *steric zipper*, where the SCs from opposing  $\beta$ -strands interdigitate to create a tightly packed, dry, and zipper-like interface. This additional stacking of the  $\beta$ -sheets introduces a second periodic spacing of approximately 10 Å, which can be observed in X-ray diffraction patterns of amyloid fibrils [8, 9].

The formation of amyloid fibrils is a complex, multi-step process that involves the aggregation of proteins into intermediate species, such as oligomers and protofibrils, before maturing into fibrils. While mature fibrils are often considered biologically inert, growing evidence indicates that oligomers formed on the pathway to fibril formation can be more toxic and play a crucial role in disease progression [10–14]. These toxic species disrupt cellular homeostasis, leading to neuronal dysfunction, cell death and ultimately cognitive decline.

The aggregation process is specific to the amyloidogenic protein involved in each particular neurodegenerative disease. In the case of AD, the primary amyloidogenic protein is the A $\beta$  peptide. This peptide is derived from the transmembrane amyloid precursor protein (APP) through sequential cleavage by  $\beta$ - and  $\gamma$ -secretase enzymes. The  $\gamma$ -secretase cleavage generates A $\beta$  peptides of varying lengths, with the 40-residue and 42-residue variants being the most prevalent. While A $\beta$  fibrils are a hallmark of AD pathology, the connection between A $\beta$  aggregation and disease progression is complex. Notably, cognitive decline in AD patients does not strongly correlate with the presence of amyloid plaques, suggesting that soluble pre-fibrillar aggregates are the primary drivers of neurotoxicity [15–19]. Supporting this hypothesis, studies have demonstrated that pore-forming A $\beta$  oligomers disrupt membrane integrity and contribute to neuronal toxicity [20, 21]. Furthermore, different A $\beta$  variants exhibit distinct aggregation behaviors, with the A $\beta$ (1-42) peptide being more prone to aggregation and considered more toxic than the A $\beta$ (1-40) variant [21–24]. These findings emphasize the importance of focusing on the early stages of A $\beta$  aggregation in order to better understand the formation and pathological role of toxic oligomers in AD. Additionally, they underscore the necessity of investigating the molecular interactions that stabilize these early aggregates and their connection to mature fibril structures.

Another neurodegenerative disorder associated with amyloid fibrils is HD, named after George Huntington who first described the disease in 1872 [25]. HD is characterized by involuntary movements, cognitive decline and psychiatric disturbances. The amyloidogenic protein involved in HD is the aptly named Huntingtin protein, which contains an elongated polyglutamine (PolyQ) stretch near its N-terminus. This PolyQ sequence is responsible for the aggregation of Huntingtin into amyloid fibrils, with the propensity for aggregation increasing as the PolyQ stretch lengthens [26–28].

The length of the PolyQ stretch is determined by variations in the IT15 gene, which encodes the Huntingtin protein. These variations arise from mutations in a stretch of trinucleotide CAG repeats that encode the glutamine (Gln) residues. During DNA replication, slippage – a process where one DNA strand misaligns and binds at an incorrect site – can lead to the insertion of additional nucleotides in the replica [29–31]. Repeating sequences, such as the CAG repeats in IT15, are particularly prone to slippage due to multiple complementary binding sites. This mechanism underlies the progressive expansion of the PolyQ stretch over generations, contributing to the hereditary nature of HD [32–34]. Importantly, the length of the PolyQ sequence is strongly correlated with the onset, severity and progression of HD. Longer PolyQ stretches lead to earlier disease onset, with a critical threshold of approximately 40 Gln residues required for the aggregation of Huntingtin into amyloid fibrils to occur [26, 28, 34, 35].

Despite significant advances, the aggregation behavior of PolyQ remains incompletely understood. Studies by Wetzel et al. have explored the critical nucleus size for PolyQ aggregation, revealing that beyond a chain length of 26 residues, the nucleus size is monomeric [27, 36, 37]. This suggests that even slightly longer chains should have a high propensity to aggregate. Furthermore, spectroscopic analyses by Crick et al. found no significant structure changes in PolyQ around the disease-associated chain-length threshold. These findings highlight the need

---

for further investigation into the folding and aggregation behavior of PolyQ as well as the molecular interactions that stabilize the aggregation nucleus, and the precise role of chain length in modulating aggregation propensity.

While the aggregation of amyloidogenic proteins is a hallmark of neurodegenerative diseases, not all amyloid fibrils are pathological. In fact, the formation of non-pathological amyloids is a crucial biological process in a variety of organisms, including bacteria, fungi, and mammals. These functional amyloids serve diverse roles, such as storage of proteins, structure, information, loss-of-function, and gain-of-function [39]. One example of a functional-amyloid-forming protein is the human parathyroid hormone (hPTH), which will only be referred to as parathyroid hormone (PTH) in this work. PTH is produced in the parathyroid glands as a pre-pro-hormone containing 115 amino acids [40]. The pre- and pro- sequences are cleaved off through proteolytic processing, which yields the mature 84-residue PTH hormone [41, 42]. This mature PTH(1-84) is stored in the secretory granules of parathyroid cells in an amyloid fibril form. Upon secretion, the fibrils disassemble rapidly into monomeric PTH, which is then released into the bloodstream, where it fulfills an important role in regulating blood calcium and phosphate homeostasis and in bone metabolism [43, 44].

In contrast to the pathological fibrils, the formation of PTH fibrils is a reversible process, with the fibril acting as a stable storage reservoir of PTH that can be rapidly mobilized when needed. This reversible behavior raises important questions about the molecular interactions that stabilize PTH fibrils and how they differ from those in pathological amyloids. Investigating the folding and aggregation behavior of PTH offers an opportunity to elucidate the thermodynamic differences between neurodegenerative and functional amyloids, shedding light onto the mechanisms that govern amyloid fibril formation and stability.

This thesis addresses these critical questions about the early stages of amyloid aggregation by investigating the dimerization of amyloidogenic proteins using a combination of the advanced Stochastic Approximation Monte Carlo (SAMC) simulation technique [45] and the Protein Intermediate Resolution Model 20 (PRIME20) [46]. By employing this powerful combination, this work aims to provide a detailed and comprehensive picture of the configuration space and the thermodynamics underlying the dimerization process of the amyloidogenic proteins PolyQ, A $\beta$  and PTH.

The PRIME20 protein model is an intermediate-resolution model. As such it sits between the atomistic all-atom models and the coarse-grained models. Atomistic models, which model proteins at the atomic level by considering every atom individually, offer a detailed and comprehensive representation of protein structure and dynamics. However, this level of detail comes at the expense of computational cost, which, significantly limits the size of systems and timescales that can be realistically simulated. On the other hand, coarse-grained models simplify protein structure and interactions by grouping multiple atoms into a single bead, often representing each amino acid as one bead. Additionally, solvent interactions are typically accounted for through mean-field approximations, eliminating the need to simulate individual solvent molecules. This reduction in complexity significantly reduces the system’s degrees of freedom, enabling simulations of larger systems and longer timescales. One-bead models have been widely and successfully applied to study protein folding and aggregation, capturing essential features of these processes [47–52]. However, the computational efficiency gained with coarse-grained models comes at the cost of reduced detail, potentially omitting crucial structural and dynamic features.

The PRIME20 model [46] bridges the gap between atomistic and one-bead coarse-grained models by achieving a balance between computational efficiency and structural detail. In the PRIME20 model, each amino acid is represented by four beads: three beads capture the backbone structure, enabling access to dihedral angles and directional hydrogen bond (HB) interactions between the backbone amino and carboxyl groups. The fourth bead represents the SC of the amino acid. The PRIME20 model is capable of differentiating between all 20 natu-

rally occurring amino acids. A clever grouping of amino acids into classes based on their SC properties reduces the number of SC interaction parameters down to just 19 from 210. This approach retains critical features of protein structures while reducing computational demands. The PRIME20 model has been successfully applied to a wide range of proteins, demonstrating its utility in studying protein folding, aggregation, and other complex behaviors [53–58].

The SAMC simulation technique [45] is a powerful Monte Carlo (MC) method [59, 60]. Unlike molecular dynamics (MD) simulations [59, 60], which solve Newton’s equations of motion to sample phase space, MC simulations use random sampling to explore configuration space. While MD seems more intuitive, as it mimics time evolution in a physical system, the random sampling approach of MC offers several advantages. Most notably, MC simulations are more efficient in exploring configuration space, making them particularly well-suited for studying complex systems. However, this gain in efficiency comes at the cost of losing dynamical information about the system. Therefore, whether MD or MC is the more suitable method depends on the specific research question that one tries to answer. The SAMC method that is used in this thesis is part of the family of flat-histogram methods. These methods aim to achieve an even visitation of energy levels, and by doing so they provide an estimation of the density of states. Since this quantity is a key factor in calculating thermodynamic properties, the SAMC method is particularly well-suited for studying the thermodynamics and structural properties of complex protein systems.

## 1.1 Research questions and aims

This thesis investigates the dimerization of amyloidogenic proteins using the SAMC simulation technique in combination with the PRIME20 protein model. The study focuses on three distinct protein dimer systems: PolyQ, A $\beta$ , and PTH. Each system presents unique challenges and opportunities for exploring their folding and aggregation behavior. While PolyQ and A $\beta$  are associated with neurodegenerative diseases, PTH represents a functional amyloid, offering a contrasting perspective on amyloid formation.

The research addresses key physicochemical questions from a computational-physical perspective. These questions are categorized into two groups: **common questions** that apply to all three protein systems, and **system-specific questions** tailored to the unique properties and behaviors of each protein system.

Common questions for all three systems:

- ▷ What are the dimer structures formed by the amyloidogenic proteins PolyQ, A $\beta$ , and PTH?
- ▷ What are the intra- and inter-molecular interactions that stabilize these dimers?
- ▷ How does the interplay of intra- and inter-molecular interactions affect the folding and aggregation process?

System-specific questions:

- ▷ How is the aggregation propensity of PolyQ influenced by chain length?
- ▷ What are thermodynamic differences between the A $\beta$ (1-40) and A $\beta$ (1-42) peptides?
- ▷ How do the aggregated dimer structures differ between A $\beta$ (1-40) and A $\beta$ (1-42)?
- ▷ How do the intra- and inter-molecular interactions in PTH dimers relate to the fibril structures?

- ▷ Do the HB networks in PTH(1-34) and PTH(1-84) dimer systems differ?
- ▷ Does the thermodynamic behavior of A $\beta$  and PTH dimers differ?
- ▷ If so, how does this relate to the differences between neurodegenerative and functional amyloids?

By applying the SAMC method and the PRIME20 model to address these research questions, this thesis aims to provide a fundamental thermodynamic understanding of these amyloidogenic systems. By focusing not only on physiological conditions but considering thermodynamic behavior over a wide temperature range, this work presents a broader perspective on the protein's aggregation behavior which is essential for understanding the molecular mechanisms underlying amyloid fibril formation.

## 1.2 Structure of the thesis

The remainder of this thesis is structured as follows: Chapter 2 provides an introduction to protein modelling in general and a comprehensive description of the PRIME20 protein model. Subsequently, Chapter 3 introduces the SAMC simulation technique and lists the relevant observables used to analyze the peptide systems. Chapters 4, 5, and 6 present the results of the dimerization simulations for the PolyQ, A $\beta$ , and PTH systems, respectively. Each chapter includes a detailed discussion of the results and their implications for the folding and aggregation behavior of the respective protein. Finally, Chapter 7 summarizes the key findings of this work and provides an outlook on future research directions.

## Chapter 2

# Protein Model

The work at hand employs computer simulations to generate data on the thermodynamic properties of protein systems. A computer simulation, in this context, consists of two components: a model representation of the system under investigation and an algorithm that defines the rules for manipulating and altering the state of the system. This chapter will focus on the model, while the algorithm, or simulation method, will be discussed in chapter 3.

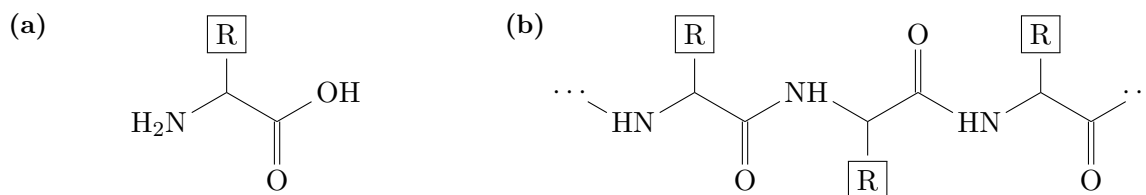
The model representation of the real polymer system is commonly referred to as the polymer model. It holds all the physical information of the system that is deemed relevant for addressing specific scientific inquiries. Typically, this information includes atom positions and sizes, as well as types of interactions between atoms and their respective energy contributions. In this study, the protein model employed is the PRIME20 [46, 53, 61]. It is a coarse-grained model which means that it simplifies the protein structure by representing multiple atoms by a single bead. In the PRIME20 model this results in four beads per amino acid.

In order to understand the modeling of proteins in general and the PRIME20 model in particular, this chapter begins with an overview of proteins and their structural properties. Subsequently, various approaches to protein modeling are discussed, encompassing considerations such as the level of detail and the selection of force fields. In this chapter's final section, the PRIME20 model is detailed, covering its geometry, interaction potentials, and the relevant parameters involved.

## 2.1 Proteins

Proteins play a vital role in life on earth. They are the most prevalent biological macromolecules and are found in all living cells where they adopt diverse roles in numerous biological functions [62]. Proteins are the building blocks for a wide array of biological structures, such as enzymes, hormones, transporters, muscle fibers and many others. Hence, in 1838 the Swedish chemist Jöns Jacob Berzelius introduced the term 'protein', derived from the Greek adjective *proteios* (πρωτειος) which signifies a position of primary importance or first rank, reflecting their significance in the biological hierarchy [63, 64].

Proteins are biological polymers. Polymers are large molecules composed of multiple repeating subunits that are of similar chemical structure. These subunits are called monomers and in the case of proteins, the monomers are the amino acids. There are 20 different amino acids that are commonly found in proteins. They all have the same backbone structure which consists of an amino group ( $\text{NH}_2$ ) and a carboxyl group ( $\text{COOH}$ ) that are bound to the same  $\alpha$ -Carbon ( $\text{C}_\alpha$ ). The general structure formula for an amino acid is shown in fig. 2.1(a). The amino acids differ from each other by their SC, also called R-group, which is attached to the  $\alpha$ -Carbon. This gives them the name  $\alpha$ -amino acids. The structure formulas of all 20 amino acid SCs are shown in fig. 2.8. The  $\alpha$ -Carbon is a chiral center, which means that it has four different substituents forming a tetrahedral structure which can exist in two different spatial



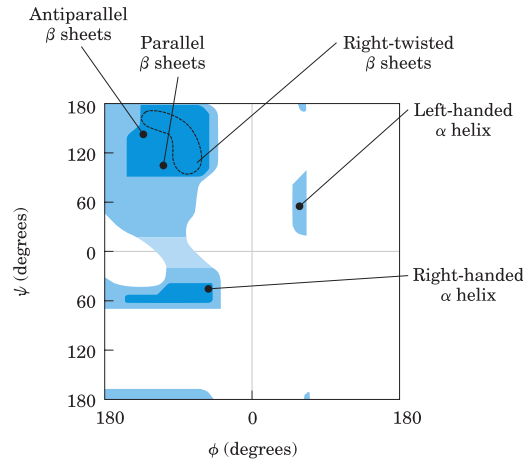
**Figure 2.1:** (a) Structure of an  $\alpha$ -amino acid and a peptide chain. The amino acid consists of an amino group ( $NH_2$ ), a carboxyl group ( $COOH$ ), and an  $\alpha$ -Carbon. The side chain ( $R$ ) is attached to the  $\alpha$ -Carbon. (b) The peptide chain is formed by linking amino acids together by peptide bonds.

arrangements. The two structures with identical chemical formula and distinct arrangement in space are called stereoisomers. They are distinguished as either 'left-handed' (Latin: *laevus*  $\rightarrow$  L-isomer) or 'right-handed' (Latin: *dexter*  $\rightarrow$  D-isomer). However, only the L-isomers of amino acids are found in protein molecules, while their D-isomers are rarely found in nature.

In order to form proteins, multiple amino acids are linked together by covalent bonds. The process involves a condensation reaction, where the carboxyl group of one amino acid reacts with the amino group of another amino acid, leading to the release of water. The resulting bond is known as a peptide bond, and the resulting chain of amino acids is called a polypeptide chain. Depending on biological function and chain length, one can distinguish between peptides and proteins. Generally, shorter chains are called peptides, while longer chains with a distinct biological function are referred to as proteins. However, since the distinction by length between the two terms is not clearly defined, and biological function does not play a significant role in this thesis, the terms *peptide*, *polypeptide*, and *protein* will be used interchangeably from here on out.

The sequence of amino acids in the polypeptide chain defines its primary structure. Additionally, there are the polypeptides secondary, tertiary and quaternary structure which describe varying levels of detail of the spacial arrangement of the peptide chain(s). The secondary structure describes the local folding of the chain, while the tertiary structure refers to the 3D structure of the entire chain. If a protein consists of multiple polypeptide chains, then the quaternary structure describes the assembly of multiple chains into a large molecular complex. The spacial arrangement of the protein chain is crucial for its functionality. Only by assuming the 'correct' structure, a protein fulfills its intended biological function.

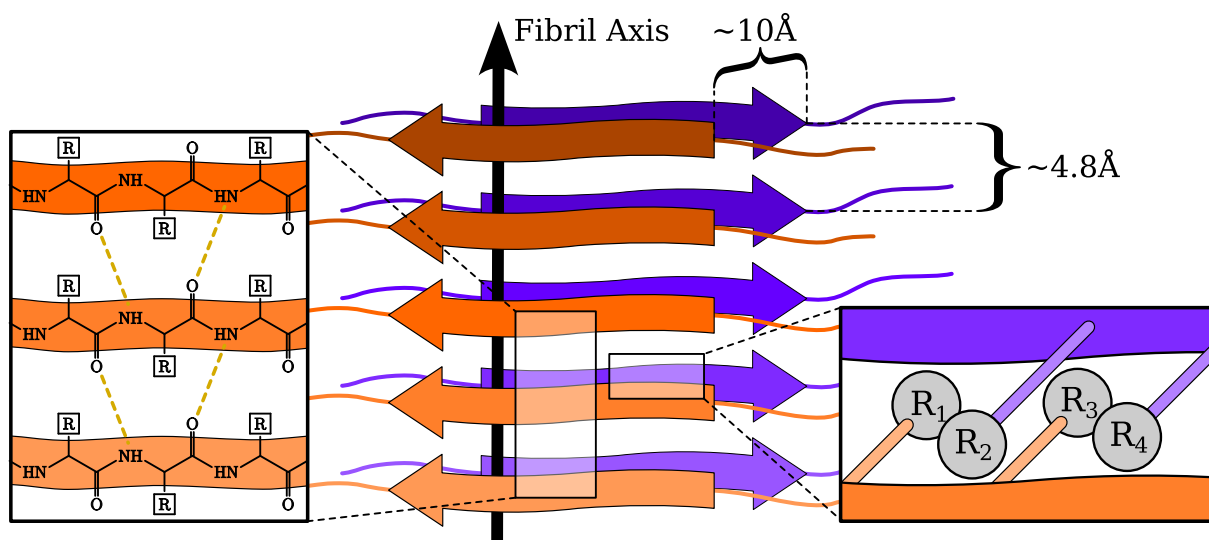
The two most common secondary structure motifs are the  $\alpha$ -helix and the  $\beta$ -sheet. They were first described by Linus Pauling and Robert Corey in 1951 [65–67]. A way of identifying both of these structures is by their backbone HB networks. In general, HBs of the protein's backbone form between the amino- and the carboxyl-group of different amino acid residues [68]. If a protein is in  $\alpha$ -helix conformation, HBs form between the amino group of the  $i^{\text{th}}$  residue and the carboxyl-group of the  $(i+4)^{\text{th}}$  residue. This is a repeating pattern and applies to all residues that are involved in the helix structure. The  $\beta$ -sheet on the other hand forms HBs between residues of adjacent stretched out chain segments. These stretched out segments are called  $\beta$ -strands. When these  $\beta$ -strands are connected by a loop, the resulting structure is known as a  $\beta$ -hairpin. This arrangement aligns two  $\beta$ -strands of the same polypeptide chain in an anti-parallel fashion and is stabilized by HBs between the  $\beta$ -strands. The classification of  $\beta$ -hairpins is based on the number of residues in the loop. The most common  $\beta$ -hairpin motifs consist of 2 or 3 residues in the loop between the first residues in the  $\beta$ -sheet that form HBs [69–71]. Accordingly, these two  $\beta$ -hairpin motifs are referred to as class 2 and class 3  $\beta$ -hairpins, respectively. In class 2  $\beta$ -hairpins, the first HB after the loop is formed between the  $i^{\text{th}}$  and  $(i+3)^{\text{th}}$  residue, thereby skipping 2 residues. In class 3  $\beta$ -hairpins, the first HB after the loop is formed between the  $i^{\text{th}}$  and  $(i+4)^{\text{th}}$  residue, thus skipping 3 residues. Unlike the repeating pattern of HBs in an  $\alpha$ -helix, the HB pattern in a  $\beta$ -hairpin is non-repeating. This occurs because in the stretched chain, neighboring residues are oriented in opposite directions (see fig. 2.1(b)). Consequently,



**Figure 2.2:** Ramachandran plot for L-Ala residues showing how different combinations of  $\phi$  and  $\psi$  angles correspond to different secondary structure motifs. The Ramachandran plots for other amino acids vary from that of Ala shown here. The blue areas designate combinations of  $\phi$  and  $\psi$  angles that are legally accessible for the protein backbone, based on calculations using known van-der-Waals radii and bond angles. Dark blue areas reflect conformations without steric overlap; medium blue areas belong to conformations only allowed at the extreme limits of unfavorable atomic contacts; and light blue areas are conformations that are only allowed if some flexibility is permitted in the bond angles. The white areas designate conformations that are not allowed due to steric hindrance. The figure was adapted from Nelson et al. [62].

subsequent HBs in the  $\beta$ -hairpin form between the next-nearest neighbors of the previous HB partners. This implies that if the initial HBs after the loop were formed between the  $i^{\text{th}}$  and  $(i + n)^{\text{th}}$  residue, then the subsequent HBs are formed between the  $(i - 2)^{\text{th}}$  and  $(i + n + 2)^{\text{th}}$  residue and subsequently between the  $(i - 4)^{\text{th}}$  and  $(i + n + 4)^{\text{th}}$  residue, and so forth. Inside the anti-parallel  $\beta$ -sheet structure, both the amino- and the carboxyl-group of one residue are forming a HB to the respective carboxyl- and amino-group of the other residue. This leads to two HBs being formed per residue contact. Next to  $\alpha$ -helices and  $\beta$ -sheets, other secondary structure motifs exist which are less common. Examples include the  $\pi$ -helix with a reoccurring HBs pattern between amino- and carboxyl-groups  $(i + 5, i)$ , the  $3_{10}$ -helix with HBs between  $(i + 3, i)$ , and the 2.27-helix or ribbon with HBs between  $(i + 2, i)$  [62, 72]. In the absence of any regular secondary structure motif, the polypeptide chain is said to be intrinsically disordered.

Along the backbone of the polymer chain exist three recurring covalent bonds: (N- $C_\alpha$ ,  $C_\alpha$ -C, C-N). The third one is the peptide bond connecting amino acid residues. Due to its partial double-bond character the peptide bond is unable to rotate freely and is planar. Its associated dihedral angle  $\omega$  can only take on values around  $\sim 180^\circ$  (*trans* configuration) or  $\sim 0^\circ$  (*cis* configuration). The other two bonds can rotate freely. Consequently, the polypeptide chain can be seen as a series of rigid planes in which consecutive planes share the  $C_\alpha$  as a common point of rotation. The rotation around the N- $C_\alpha$  and the  $C_\alpha$ -C bonds are described by the dihedral angles  $\phi$  and  $\psi$ , respectively. Equivalent to the HB networks described above, these dihedral angles are important parameters to characterize the secondary structure of proteins. A common method are Ramachandran plots [73–76]. In a Ramachandran plot the  $\phi$  and  $\psi$  angles of a residue are represented as a point in a 2D map. Residues that are part of different secondary structure motifs occupy different regions in the Ramachandran plot. How the different regions correspond to the different secondary structure motifs is shown in fig. 2.2. The Ramachandran plot is a powerful tool to analyze the secondary structure of proteins and has been used repeatedly in their identification.



**Figure 2.3:** Schematic representation of the cross- $\beta$ -sheet motif in amyloid fibrils. The  $\beta$ -strands are oriented perpendicular to the fibril axis. The  $\beta$ -strands are connected by HBs between the backbone atoms. The two stacked  $\beta$ -sheets are held together by hydrophobic interactions between the side chains in a steric zipper.

### 2.1.1 Amyloid fibrils

A common feature of the proteins investigated in this work is their ability to form amyloid fibrils. Amyloid fibrils are highly ordered protein aggregates that are associated with a variety of diseases, including Alzheimer’s disease and Huntington’s disease, but also with non-pathological functions such as storage of hormones like PTH (see chapter 1). They are super-molecular fibrillar structures that can reach up to micrometer lengths. Within the amyloid fibril, the defining structural motif is the cross- $\beta$  structure, which is shown schematically in fig. 2.3.

The cross- $\beta$  motif consists  $\beta$ -strands that are oriented perpendicular to the fibril axis. These  $\beta$ -strands are connected by HBs between the backbone atoms to form a  $\beta$ -sheet, which results in the characteristic inter-strand spacing of approximately 4.8 Å. The  $\beta$ -strands within the sheets can adopt either an **anti-parallel** orientation, as observed in PolyQ fibrils [77], or a **parallel** orientation, as seen in fibrils formed by A $\beta$  [78] and PTH [79]. An example of the parallel arrangement is depicted in fig. 2.3, where broad arrows represent  $\beta$ -strands, with  $\beta$ -strands of the same color (orange or purple) belong to the same  $\beta$ -sheet. The zoom-in graphic on the left side of fig. 2.3 portrays the hydrogen bonding network within a  $\beta$ -sheet.

In the fibril, two  $\beta$ -sheets are paired, and the structure is stabilized by a closely packed dry interface of self-complementary side chains protruding from the  $\beta$ -sheets. This is illustrated by the zoom-in graphic on the right of fig. 2.3. This interface is known as a *steric zipper*. The steric zipper creates an inter-sheet spacing of approximately 10 Å, a key feature consistently observed in X-ray diffraction patterns of amyloid fibrils [3, 8, 9, 78–80].

## 2.2 Modeling of proteins

When studying polypeptides one needs to choose an appropriate model. The choice of model depends on the system one wishes to study and the scientific question that one wishes to answer. Broadly speaking, protein models fall into one of two categories: atomistic or coarse-grained (CG) models. The two categories differ in the level of detail they provide. This allows for a trade-off between the accessibility of certain information about the system and the computational cost of the simulation.

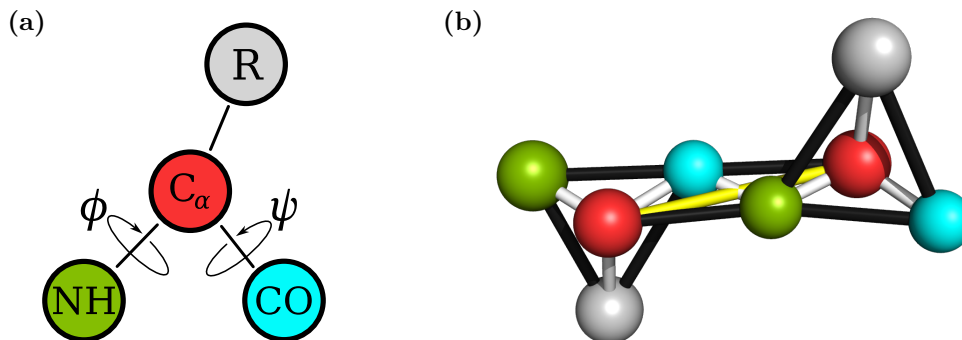
**Atomistic models** represent the arguably most intuitive approach to protein modeling by representing each atom in the system individually. This granular representation allows for a com-

prehensive description of the system and the interactions between its constituents. Interactions among atoms are typically governed by a force field, which is a set of equations and parameters designed to reproduce the geometry, dynamics, thermodynamics, and other chemical and physical properties of a system. Despite their thoroughness, force fields are always an abstraction of the underlying electronic structure of the atoms, typically derived from experimental data and quantum mechanical calculations. Models that explicitly consider quantum mechanical effects are known as *ab initio* models. While atomistic models excel in providing detailed insights, this level of granularity comes at high computational cost and time effort. A protein system for example may easily contain hundreds of atoms. In addition, solvent molecules have to be considered which further increases the number of atoms in the system. Updating the positions and interactions between such a huge number of particles renders calculations on atomistic models computationally expensive. This imposes limitations on time scales and system sizes that can be effectively studied.

**Coarse-grained models** take a different approach by grouping atoms together and representing them as a single bead. This simplification significantly reduces the number of particles in the system, leading to more efficient calculations. Consequently, CG models prove invaluable for investigating large systems and long time scales. However, this simplification comes at the cost of losing detailed information about the system. By condensing multiple atoms into a single bead, potentially crucial information is neglected, resulting in a decreased level of detail compared to atomistic models. Nevertheless, CG models are capable of reproducing key structural and thermodynamic properties of the system and have been successfully applied in numerous studies of polymer systems [48]. The level of detail a CG model can provide varies significantly depending on its degree of coarse-graining. Some CG models are designed to be generic and applicable across various polymer systems. Classical examples of such CG polymer models include the bead-spring model [81, 82] and the lattice models [83]. On the other hand, some CG models are tailored specifically towards representing proteins, and they include certain structural details that are specific to particular proteins [84].

Another important distinction between types of CG models is made by their method of establishing the force field. On the one hand, physics-based models derive their interaction potentials from the inherent physical properties of the system’s constituents [85–88]. On the other hand, there are the knowledge-based models which construct their force fields based on known secondary and tertiary peptide structures. This approach renders knowledge-based models particularly useful for simulating the native structures of proteins but less effective at modeling their disordered states. A prominent example of knowledge-based models is the Gō-model [89, 90] which, along with its derived models [91–93], has been widely utilized in numerous studies. The core concept of the Gō-model revolves around capturing the protein’s native globular state and then defining energy potentials which primarily influence beads that are in close proximity when the protein is in its native state. Obviously, this puts a bias of the model towards the proteins native state.

As mentioned before, CG models vary in their level of detail. A popular approach is to represent each amino acid by a single bead [47, 49, 50, 52, 94]. However, this approach denies access to information about the internal structure of the amino acids in the chain. An intermediate-resolution approach between one-bead representation and atomistic modeling are the four-bead models. In these models, the backbone of the protein is typically represented by three beads [95, 96], which allows for access to the dihedral angles  $\psi$  and  $\phi$  of the backbone. The SC is then modeled by a single bead. This simplification reduces the number of particles in the system while still providing access to important structural information. An example of a four-bead model is the PRIME20 model [46] which was used in this study and will be discussed in the following section.



**Figure 2.4:** (a) Schematic representation of an  $\alpha$ -amino acid residue in the PRIME20 model. The backbone of the amino acid residues is represented by three beads: the green bead represents the amino group (NH), the red bead the  $\alpha$ -Carbon ( $C_\alpha$ ), and the cyan bead the carboxyl group (CO). The side chain is represented by a single gray bead. In (b) two amino acid residues are connected by a peptide bond. Physical bonds between the beads are depicted as white sticks, while pseudo-bonds are shown in black and yellow.

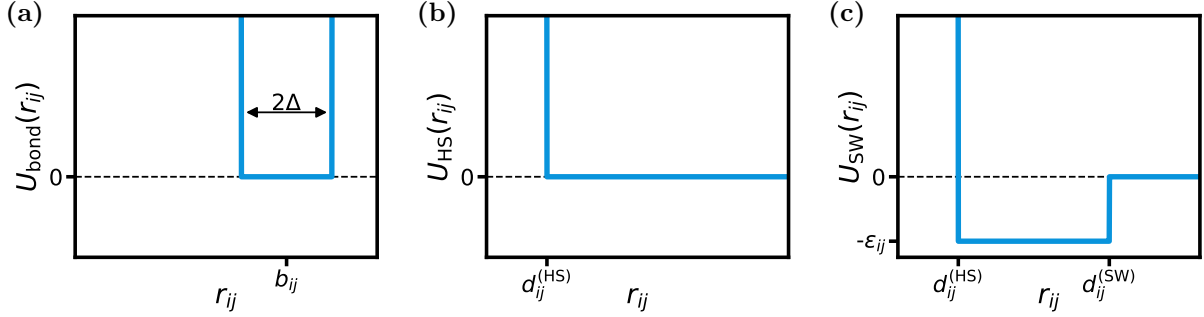
## 2.3 PRIME20 protein model

The Protein Intermediate Resolution Model 20 (PRIME20) model is a four-bead CG model for proteins, designed to study their secondary structure formation. It is an extension of the Protein Intermediate Resolution Model (PRIME) model which was originally developed by Voegler Smith et al. [97–100] to investigate the secondary structure formation of polyalanine. The PRIME20 model, developed by Cheon et al. in 2010, expands the PRIME model to include all 20 amino acids [46]. A significant achievement of the PRIME20 model is a reduction of the number of energy parameters required to describe residue interactions, which was realized through strategic grouping of SCs (see 2.3.1). This approach mitigates some of the bias towards native structures inherent to knowledge based models as discussed previously. Both model variants utilize discontinuous interaction potentials, which allowed for simulations using an efficient molecular dynamics (MD) algorithm called discontinuous molecular dynamics (DMD) [97]. Although it is not a requirement, discontinuous potentials also make the PRIME20 model a well suited candidate for MC simulations. Additionally, the four-bead representation allows for a more detailed modeling of backbone HBs by introducing directional HB potentials, enhancing the model’s ability to capture native protein structures.

The PRIME and PRIME20 model share identical basic structures and backbone geometry. They differ from each other in the SC parameters and interactions. In the subsequent discussion, the geometry, interaction types, and model development, including the reduction of the number of energy parameters, will be explored in detail. Furthermore, a comprehensive overview of the model’s parameters relevant to the systems investigated in this work will be provided.

### 2.3.1 Geometry and interaction potentials

As mentioned earlier, the PRIME20 model employs a four-bead representation for the amino acid residues. This is illustrated in fig. 2.4(a). Three beads are used to model the backbone of the amino acids: the green bead represents the amino group (NH), the red bead the  $\alpha$ -Carbon ( $C_\alpha$ ), and the cyan bead the carboxyl group (CO). The fourth bead in gray (R) represents the SC of the amino acid. The backbone beads are centered at the position of their respective N and C atoms and the R-bead is placed at the center of mass position of the SC. With this assignment of atoms to the beads, the PRIME20 model has a relatively detailed backbone representation, which allows access to the dihedral angles  $\phi$  and  $\psi$ . However, the single-bead representation of the SC is a rather coarse approximation, particularly for complex SCs like those of arginine or



**Figure 2.5:** Schematic representation of the interaction potentials in the PRIME20 model. (a) Bond potential. (b) Hard-sphere potential. (c) Square-well potential.

tryptophan.

Interactions between bonded beads are modeled by infinite well potentials as shown schematically in fig. 2.5(a). This allows the bond lengths to fluctuate around an ideal bond value:

$$U_{\text{bond}}(r_{ij}) = \begin{cases} 0 & \text{if } (1 - \Delta)b_{ij} \leq r_{ij} \leq (1 + \Delta)b_{ij} \\ \infty & \text{else,} \end{cases} \quad (2.1)$$

where  $r_{ij}$  is the distance between beads  $i$  and  $j$ ,  $b_{ij}$  is the ideal bond length, and  $\Delta$  is the fluctuation parameter. The fluctuation parameter is set to  $\Delta = 2.375\%$  [101]. The value of  $b_{ij}$  is specific to the interacting bead types and will be discussed in section 2.3.2 and appendix A. In addition to the physical bonds between neighboring beads (NH-C $_{\alpha}$ , C $_{\alpha}$ -CO, CO-NH and C $_{\alpha}$ -R), the model also includes pseudo-bonds. They are treated like the physical bonds, according to eq. (2.1). Figure 2.4(b) shows two amino acid residues represented in the PRIME20 model that are connected by the peptide bond. Physical bonds are depicted as white sticks, while pseudo-bonds are shown in black and yellow. Generally, the pseudo-bonds in PRIME20 eliminate the need for bond angle potentials in the model. Those depicted as black sticks are established between beads that are next-nearest neighbors and which are thus separated by 2 regular bonds. The pseudo-bonds involving the SC bead R ensure that the residue remains in L-isomerization. The pseudo-bond depicted as a yellow stick in fig. 2.4(b) is established between consecutive C $_{\alpha}$  beads, maintaining a distance of 3.8 Å between these atoms. This setup ensures the polymer remains in *trans* conformation. The *cis* variant of the peptide bond is not considered in the PRIME20 model.

Non-bonded beads generally interact via hard-sphere repulsion (see fig. 2.5(b)). This assigns a specific hard-sphere diameter  $d_i^{(\text{HS})}$  to each bead  $i$ . When two beads  $i$  and  $j$  interact, the effective bead diameter  $d_{ij}^{(\text{HS})}$  is calculated as the arithmetic mean using the Lorentz-Berthelot combining rule [102]:

$$d_{ij}^{(\text{HS})} = \frac{1}{2} \left( d_i^{(\text{HS})} + d_j^{(\text{HS})} \right). \quad (2.2)$$

However, there are exceptions to the regular treatment of hard-sphere repulsion in ten specific instances where beads are separated by 3, 4, or, in one case, 6 bonds along the chain. In these instances, specific *squeeze parameters* are applied, altering their effective bead diameters. Further details on these parameters will be provided in section 2.3.2. It is important to note that bead diameters vary depending on the context, so the bead diameters shown in fig. 2.4 are merely illustrative and were chosen with the purpose of visual clarity.

In addition to hard-sphere repulsion, there are two types of non-bonded beads that experience additional interactions: 'side-chain – side-chain' and the backbone HB interactions. Attractive interactions are modeled using square-well potentials, while repulsive interactions are modeled using square-shoulder potentials (see fig. 2.5(c)). Both types of 'side-chain – side-chain'

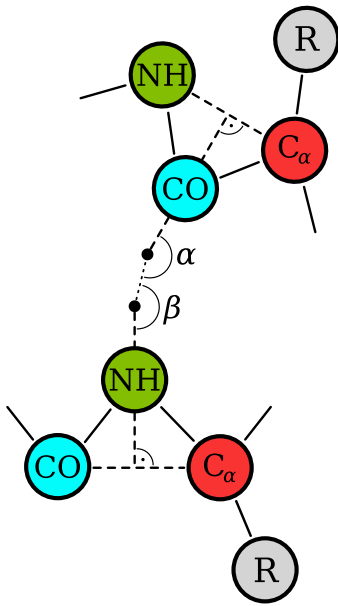
interactions depend solely on the distance between the interacting beads, with the attractive and repulsive interactions being described by the same equation:

$$U_{\text{sw}}(r_{ij}) = \begin{cases} \infty & \text{if } r_{ij} < d_{ij}^{(\text{HS})} \\ \varepsilon_{ij} & \text{if } d_{ij}^{(\text{HS})} \leq r_{ij} < d_{ij}^{(\text{SW})} \\ 0 & \text{if } d_{ij}^{(\text{SW})} \leq r_{ij} \end{cases} \quad (2.3)$$

where  $r_{ij}$  is the distance between beads  $i$  and  $j$ ,  $d_{ij}^{(\text{HS})}$  is the hard-sphere diameter,  $d_{ij}^{(\text{SW})}$  is the square-well distance, and  $\varepsilon_{ij}$  is the interaction strength. The last three parameters are specific to the interacting bead types. The interaction strength  $\varepsilon_{ij}$  can be either negative or positive, which results in an attractive or repulsive interaction, respectively.

Backbone HBs form between the NH and the CO beads. They are modeled with square-well potentials (eq. (2.3)), akin to the 'side-chain – side-chain' interactions. However, HB interactions introduce an additional constraint related to the angle between the HB and bonds in the backbone of the peptide chain. In real proteins, a HBs are most stable when they exhibit coaxial geometry, where the N-H-O and the C-O-H angles are both  $180^\circ$  and the H and the O atoms are positioned  $2 \text{ \AA}$  apart. However, deviations of the angle are common, with angles down to  $110^\circ$  still being considered indicative of HB formation [68]. In the PRIME20 model, four criteria have to be met for a HB to form:

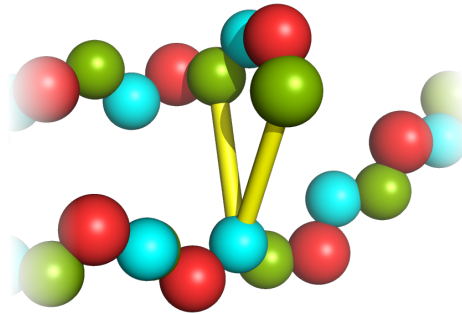
1. Neither of the two beads are part of an established hydrogen bond.
2. The residues associated with the two beads are separated by at least 3 other residues.
3. The NH and CO beads are within the HB square-well distance. In the original PRIME model, this distance was set to  $d_{ij}^{(\text{SW})} = 4.2 \text{ \AA}$  [97]. In the PRIME20 model, this distance was increased to  $d_{ij}^{(\text{SW})} = 4.5 \text{ \AA}$  [46] which provides a more accurate description of HB formation in proteins with larger SCs.
4. Both two angles  $\angle(\text{N-H-O})$  and  $\angle(\text{C-O-H})$  fall within the range of  $[120^\circ, 180^\circ]$ .



**Figure 2.6:** Schematic representation of the construction of the virtual positions of the amide hydrogen and the carboxyl oxygen, used for hydrogen bonding.

The implementation of the fourth criterion demands some further explanation since neither the amide hydrogen nor the carboxyl oxygen are modelled explicitly in the PRIME20 model. Instead, virtual positions of the atoms are calculated. Therefore, lines are constructed that go through the respective N and C atoms, and are perpendicular to the pseudo-bond between the neighboring beads. The virtual position of the hydrogen atom is then placed at a distance of  $1 \text{ \AA}$  from the N atom along the constructed line. Likewise, the location of the oxygen atom is at a distance of  $1.2 \text{ \AA}$  from the C atom along its line. With this construction, the angles  $\angle(\text{N-H-O})$  and  $\angle(\text{C-O-H})$  can be calculated and their adherence to criterion 4 can be checked. The method is illustrated in fig. 2.6, where the dots at the end of the dashed lines symbolize the virtual N and O atoms. The dashed lines themselves represent the constructed lines through the N and C atoms and the angles  $\alpha$  and  $\beta$  represent the angles  $\angle(\text{C-O-H})$  and  $\angle(\text{N-H-O})$ , respectively.

In the original PRIME model, an additional method was used to stabilize the HB once they were formed [97]. This method involved the introduction of auxiliary interactions between the NH and CO beads and the beads that neighbor



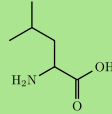
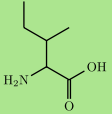
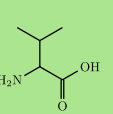
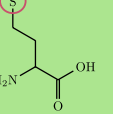
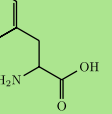
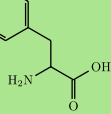
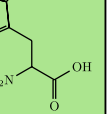
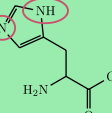
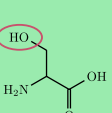
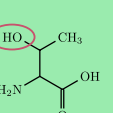
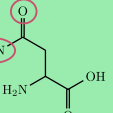
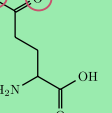
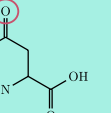
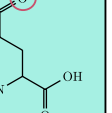
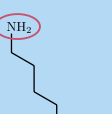
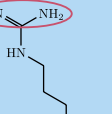
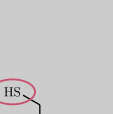

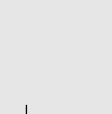
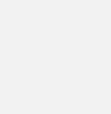
**Figure 2.7:** Schematic representation of a peptide conformation in the PRIME20 model, where a single CO (cyan) bead has two possible NH (green) beads to form a hydrogen bond (yellow sticks) with. Side chain beads (R) are hidden for visual clarity.

their HB partners. Once a HB is formed, the effective hard-sphere diameters  $d_{ij}^{(\text{HS})}$  between the NH bead and both neighbors of the CO bead is increased, as well as  $d_{ij}^{(\text{HS})}$  between the CO bead and both neighbors of the NH bead. In DMD simulations for which the PRIME20 model was originally designed for, this effectively reduces the risk of breakage for established HBs. However, since dynamic bending processes like those found in molecular dynamics simulations do not exist in MC simulations, the auxiliary interactions effectively serve the same purpose as the fourth criterion for HB formation. They were implemented as an additional constraint on HB formation, regardless.

Another criterion for HB formation that warrants further discussion is the first one, which requires that neither of the two considered beads can already be part of an existing HB. This criterion acknowledges the possibility of a peptide conformation where a single bead can participate in two different HBs. In other words, one bead can serve as a partner in two separate HBs, allowing both to form simultaneously. A schematic representation of such a conformation is depicted in fig. 2.7. The first criterion then makes sense when considering the dynamics of the polymer system in 'real' time as is the case for DMD simulations, the PRIME20 model was originally designed for. In a polypeptide system evolving in time, the partial positive charge of the donor group (NH) is already engaged in a HB and compensated for by the negatively charged acceptor group (CO). Thus, it is unlikely for the proton to engage in another HB. However, in MC simulations, there is no 'real'-time development of the system. If multiple HB are possible, none of them should be considered preexisting and more favorable than the other. Moreover, MC moves allow for large scale changes in the system in a single step. These moves can potentially result in a conformation where two HB involving the same bead can newly form without one already existing. Since none of the HBs are already established, the question arises of how the PRIME20 model determines which HB to form. This question is solved by the SAMC simulations requirement of a unique and unambiguous assignment of configurations to energies. Thus, a random selection of HB would violate this requirement. Instead, the PRIME20 checks for HB formation by sequentially evaluating residues from the N-terminus to the C-terminus. If a potential HB meeting all criteria is found, it is formed. If multiple HB are possible, the first one encountered is the one that is formed. This method ensures that the HB formation is unambiguous and unique. However, it introduces a bias towards HBs formation at the N-terminus of the chain. This bias is not present in real proteins and is a limitation of the PRIME20 model.

### Reduced units

In the PRIME20 model a HB yields an energy gain of  $E_{\text{HB}} = -1$ . This sets the energy scale for the entire model. As a result, all energy related physical quantities are given relative to the energy of a HB interaction, denoted as  $\varepsilon_{\text{HB}}$ . By assigning a value to  $\varepsilon_{\text{HB}}$  one can map the reduced energy  $E$  and temperature  $T$  of the model to physical quantities  $E' = \varepsilon_{\text{HB}} E$  and

Leucine	Isoleucine	Valine	Methionine	Phenylalanine	Tyrosine	Tryptophan
						
Small Hydrophobic				Large Hydrophobic		
Histidine	Serine	Threonine	Asparagine	Glutamine	Aspartic Acid	Glutamic Acid
						
Polar					Negative Charge	
Lysine	Arginin	Cysteine	Proline	Alanine	Glycine	
						
Positive Charge						

**Figure 2.8:** Classification of amino acids in the PRIME20 model. The Amino acids are represented by their structure formula, their 3- and their 1-letter name. They are grouped based on their side chain properties. Groups and atoms that are able to form side chain hydrogen bonds are marked by purple circles. The figure is adapted from the PhD thesis of Böker (2019) [61].

$T' = \varepsilon_{\text{HB}} T / k_B$ , respectively. The PRIME20 model utilizes an implicit solvent, implementing a mean field approximation for its potentials. This approach does not explicitly simulate solvent molecules and solvent interactions. Instead, it incorporates their effects by averaging them into the overall energy scale. As a result,  $\varepsilon_{\text{HB}}$  must account for the energy gain of backbone HBs, as well as other interactions such as those between peptide and solvent, and potential side chain hydrogen bonds. The specific value of  $\varepsilon_{\text{HB}}$  is unique to the modelled peptide system and must be determined for every investigated system individually. For PolyQ, see section 4.3; for A $\beta$ , see section 5.1.1; for PTH, see section 6.1.1.

## Model development

The PRIME20 model was developed by Cheon et al. in 2010 [46]. The authors used an energy gap optimization approach to derive the model's interaction parameters. The approach assumes that the native state of a protein resides in the global minimum of the free energy landscape [103]. Thus, the goal of the optimization was to identify interaction parameters that make the energy of the native state lower than the energies of a large number of decoy structures. Therefore, a set of 711 structures from the protein data bank (PDB) in their native states was compared to nearly 2 million decoy structures generated using a gapless threading algorithm [104]. The optimization of the energy parameters was performed using a modified stochastic perceptron learning algorithm [46, 105]. Additionally, radial distribution functions for all pairs of united atoms were calculated for the native structures to estimate the effective hard sphere diameters.

What distinguishes the PRIME20 model from other knowledge-based optimization ap-

proaches is its significant reduction in the number of independent pair interaction parameters. The authors aimed to strike a balance between the number of parameters and the accuracy of the model. Traditional optimization approaches for interactions among 20 amino acid types typically result in  $\sum_{n=1}^{20} n = 210$  energy parameters [106, 107]. In contrast, the PRIME20 model reduced this number to just 19 pair interaction parameters. This reduction was achieved by grouping amino acids according to their SC properties. As illustrated in fig. 2.8, the amino acids were classified into 14 groups based on properties such as hydrophobicity, polarity, size, charge, and HB donor/acceptor properties. The four amino acids cysteine (Cys), proline (Pro), alanine (Ala), glycine (Gly) were treated as special cases. By assuming that Gly has no SC and by considering other specific properties, the initial grouping resulted in  $\sum_{n=1}^{13} n = 91$  energy parameters. These were further reduced by combining similar interaction types. For example, large hydrophobic SCs (phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp)) were combined with small hydrophobic SCs (leucine (Leu), isoleucine (Ile), valine (Val)) because the latter do not form SC HBs and the polar groups of Tyr and Trp are not relevant. This reduction ultimately resulted in a total of 19 energy parameters for SC interactions [46]. The exact values of these parameters will be discussed and presented for a selection of amino acids in the following section. A complete list of the parameters of the PRIME20 model can be found in appendix A.

### 2.3.2 Parameters

The parameters of PRIME and PRIME20 can be found in numerous publications (PRIME: [108–111]; PRIME20: [46, 54, 55, 58, 112–119]). However, until recently there was no single source that contained all the parameters as a complete set. Thankfully, Böker et al. undertook the task of reviewing all available literature to compile a comprehensive list of parameters for the PRIME20 model, which they published in [53]. The lists of parameters presented here are derived from their research and have been implemented in this work.

**Table 2.1:** List of ideal bond lengths ( $b_{ij}$ ), ideal pseudo-bond lengths ( $b_{ij}^{(\text{pseudo})}$ ), bead diameters ( $d_{ij}^{(\text{HS})}$ ) and bond angles of the backbone of the protein chain in the Prime model. Lengths and diameters are given in Å, angles in rad.

$b_{ij}$		$b_{ij}^{(\text{pseudo})}$		$d_{ij}^{(\text{HS})}$		Bond angle	
NH–C $_{\alpha}$	1.46	NH $_i$ –CO $_{i+1}$	4.25	NH	3.3	$\angle(\text{NH}–\text{C}_{\alpha}–\text{CO})$	1.937
C $_{\alpha}$ –CO	1.51	C $_{\alpha i}$ –NH $_{i+1}$	2.41	C $_{\alpha}$	3.7	$\angle(\text{C}_{\alpha}–\text{CO}–\text{NH})$	2.025
CO–NH	1.33	CO $_i$ –C $_{\alpha i+1}$	2.45	CO	4.0	$\angle(\text{CO}–\text{NH}–\text{C}_{\alpha})$	2.129
		C $_{\alpha i}$ –C $_{\alpha i+1}$	2.80				

We will first look at the parameters of the backbone geometry since it is identical for the different amino acids, and therefore as well for the two model variants PRIME and the PRIME20. In table 2.1 the bond lengths, pseudo-bond lengths, bead diameters, and bond angles of the polymer backbone in the PRIME20 model are listed. The lengths are given in Å, and the bond angles in rad. The bond lengths are the ideal values around which the distance between bonded beads can fluctuate by  $\Delta = 2.375\%$  [53, 101] (see eq. (2.1)). In the 2001 paper by Voegler Smith et al. [97] which first introduced the PRIME model, the originally chosen square-well diameter of 4.2 Å for HB interactions can be found. This was later increased to  $d_{ij}^{(\text{SW})} = 4.5$  Å [46] for the PRIME20 model, which is the value used in this work.

Backbone interactions also include HB formation. Once a HB is formed, auxiliary interactions stabilize the HB by increasing the effective hard-sphere diameters between the specific beads involved in the HB and their neighbors. This has been already discussed in a previous section. The parameters for the auxiliary interactions of HBs are listed in table 2.2.

In extending the PRIME model to the PRIME20 model, individual parameters were assigned to interactions between different amino acids. Since the backbone remains unchanged, these

**Table 2.2:** Distances for auxiliary interactions after hydrogen bond formation between beads  $\text{NH}_i$  and  $\text{CO}_j$ . All distances are given in Å.

$\text{NH}_i\text{--C}_{\alpha j}$	$\text{CO}_j\text{--CO}_{i-1}$	$\text{NH}_i\text{--NH}_{j+1}$	$\text{CO}_j\text{--C}_{\alpha i}$
5.00	4.86	4.74	4.83

adjustments only concerns SC interactions. The protein chains investigated in this work are PolyQ, A $\beta$  and PTH. Together, they contain 19 of the 20 amino acids found in nature — all except Cys. As PolyQ is a homopolymer chain and consists only of Gln, glutamine, along with the three most commonly occurring residue types in A $\beta$  and PTH <sup>(2)</sup>, will serve as example residues in the following discussion. A full list of the parameters for all 20 amino acids and their pairings is available in the appendix of this work (see appendix A).

**Table 2.3:** Bond and pseudo-bond lengths between the selected side chains and their backbone beads. Furthermore, the masses of the side chain ( $\text{Mass}^{(\text{R})}$ ) and backbone beads ( $\text{Mass}^{(\text{BB})}$ ) are listed. All lengths are given in Å and all masses in Dalton (Da).

	$\text{C}_{\alpha}\text{--R}$	$\text{NH--R}$	$\text{CO--R}$	$\text{Mass}^{(\text{R})}$	$\text{Mass}^{(\text{BB})}$
Gln	3.300	3.750	4.000	72.0	NH 15.0
Val	2.002	2.775	2.959	43.0	$\text{C}_{\alpha}$ 13.0
Leu	2.625	3.299	3.500	57.0	CO 28.0
Lys	3.550	4.050	4.250	73.0	

The ideal bond and pseudo-bond lengths, as well as the masses of the SC beads for Gln, Val, Leu, and lysine (Lys), are listed in table 2.3. The lengths are provided in Angstrom (Å) and the masses in Dalton (Da). Additionally, the masses of the backbone beads are included for reference. The masses of the SC beads are determined by summing the masses of the atoms represented by each bead, following the same method used to determine the masses of the backbone beads.

**Table 2.4:** Energies (square-well depths) for interactions between the selected amino acids. All energies are given in units of  $\varepsilon_{\text{HB}}$ 

	Energy ( $\varepsilon_{\text{HB}}$ )			
	Gln	Val	Leu	Lys
Gln	-0.080	0.015	0.015	-0.086
Val		-0.200	-0.200	0.015
Leu			-0.200	0.015
Lys				0.073

SC interactions are modeled by square-well potentials (eq. (2.3)). The interaction strengths  $\varepsilon_{ij}$  are given in units of  $\varepsilon_{\text{HB}}$ . The interaction range (square-well widths)  $d_{ij}^{(\text{SW})}$  and the hard-sphere diameters  $d_{ij}^{(\text{HS})}$  are given in Å. The energy parameters for the selected amino acids are listed in table 2.4. The diameters and interaction ranges are listed in table 2.5. The bead diameters as well as the square-well width and depth vary for interactions between different amino acids. The complete set of parameters can be found in the appendix of this work (see appendix A). For interactions between SC beads and backbone beads, the self-interaction bead diameter is used for the SC bead when calculating the effective bead diameters of the hard-sphere interaction (eq. (2.2)).

<sup>(2)</sup>For the amino acid sequences of A $\beta$  and PTH see eqs. (3.31) and (3.32)

**Table 2.5:** Side chain bead diameters and square-well widths for interactions between the selected amino acids. All lengths are given in Å

	Bead diameter (Å)				Square-well width (Å)			
	Gln	Val	Leu	Lys	Gln	Val	Leu	Lys
Gln	3.6	3.3	3.5	3.4	6.6	6.5	6.3	6.7
Val		3.3	3.0	3.1		6.3	6.2	6.6
Leu			3.4	3.5			6.4	6.5
Lys				3.5				6.9

As Böker et al. have shown [53], the thus far here presented set of parameters fails to reproduce peptide configurations that are to be expected in realistic polypeptides. The missing features are the so called '*squeeze parameters*'. These parameters are used to adjust the effective bead diameters of beads that are close to each other along the chain and will be discussed below.

### Squeeze parameters

In the original publication of the PRIME model [97], it is mentioned that for interactions between pairs of beads separated by three or fewer bonds, an overlap of the beads by up to 25% is allowed. This corresponds to a reduction of the effective bead diameter to 75% of the regular diameter. The goal of the reduction was to allow for physically reasonable motion around the NH-C $_{\alpha}$  and C $_{\alpha}$ -CO bonds. However, as mentioned above, the model fails to reproduce certain realistic polypeptide conformation. As shown by Böker et al. [53], the model as it is presented thus far produces  $\gamma$ -helices as a folded state for polyalanine chains, which is an unrealistic result, since the folded state of polyalanine is known to be the  $\alpha$ -helix. At least one additional parameter reducing the effective bead diameters for 4-bond distances is necessary for this kind of secondary structure to form. To this end, the PRIME20 model applies 10 squeeze parameters to specific bead interactions.

**Table 2.6:** Squeeze parameters *sqz1–5* for pure backbone interactions and how they affect the effective bead diameters.

	<i>sqz1</i>	<i>sqz2</i>	<i>sqz3</i>	<i>sqz4</i>	<i>sqz5</i>
	C $_{\alpha i}$ -CO $_{i+1}$	NH $_{i-1}$ -C $_{\alpha i}$	CO $_i$ -NH $_{i+2}$	NH $_i$ -NH $_{i+1}$	CO $_i$ -CO $_{i+1}$
<i>sqz</i>	1.1436	0.88	0.87829	0.8	0.7713
'Old' $d_{ij}^{(HS)}$ (Å)	3.85	3.5	3.65	3.3	4.0
'New' $d_{ij}^{(HS)}$ (Å)	4.40286	3.08	3.2057585	2.64	3.0852

The squeeze parameters (*sqz*) can be divided into two groups of five. The first five parameters (*sqz1–5*) are applied to interactions between backbone beads. They can also be found in the code appended to the PhD thesis of Voegler Smith [100, p. 256], however they were not mentioned in the thesis text itself. The parameters *sqz1–5* are listed in table 2.6. They affect bead interactions of 3-bond distance (*sqz4*, *sqz5*), as well as some 4-bond distance interactions (*sqz1–3*). The second group of five parameters (*sqz6–10*) are applied to interactions between SC beads and backbone beads. They can also be found in the code appended to Voegler Smith [100, p. 256]. They affect 3-bond (*sqz7*, *sqz8*), 4-bond (*sqz6*, *sqz9*), and 6-bond (*sqz10*) distance interactions. The parameters *sqz6–10* are listed in table 2.7. A full list of the *sqz* parameters can again be found in table A.5.

**Table 2.7:** Squeeze parameters  $sqz6$ – $10$  for interactions between side chains and backbone. Also, the resulting diameters for relevant side chains (values are in Å).

	$sqz6$	$sqz7$	$sqz8$	$sqz9$	$sqz10$
	$R_i\text{--CO}_{i-1}$	$R_i\text{--NH}_{i+1}$	$R_i\text{--C}_{\alpha i+1}$	$R_i\text{--C}_{\alpha i-1}$	$R_i\text{--C}_{\alpha i-2}$
$sqz$	0.7607	0.7930	1.0956	1.1244	0.9259
Gln	4.139	3.996	5.062	5.134	5.000
Val	3.312	3.000	4.353	4.598	4.997
Leu	3.918	3.724	4.863	4.936	5.001
Lys	4.384	4.191	5.163	5.323	4.974

## Chapter 3

# Simulation Method

In my work, I used the advanced flat-histogram Monte Carlo method Stochastic Approximation Monte Carlo (SAMC) to study the dimerization of different proteins in the Prime20 model. The latter has been discussed in the previous chapter 2. This chapter will give an introduction to the basic concepts of statistical mechanics and the Monte Carlo method, followed by a description of the SAMC method including the system specific program features and parameters relevant to the investigated systems.

### 3.1 Statistical mechanics

In order to understand the Monte Carlo (MC) simulation method used in this work, it is necessary to introduce some basic concepts of statistical mechanics. This section will give a brief overview of statistical concepts and methods deemed important to this work and in particular for understanding the following sections on the MC method. For a comprehensive introduction to statistical mechanics, the reader is encouraged to consult the standard literature such as [120] and [121].

#### 3.1.1 Statistical ensembles

Statistical mechanics becomes relevant, when describing macroscopic systems. These are systems that consist of a very large number of particles, like atoms or molecules. If one would use a classical mechanics description, one would need to establish and integrate the equations of motion. Because the number of these equations is equal to the degrees of freedom in the system, for systems with a large number of particles this quickly becomes impracticable. And even if one would manage to set up all the equations of motion there would still remain the task of determining the initial conditions for the positions and velocities of all the particles. When considering for example a room filled with approximately  $N_A \approx 6.022 \cdot 10^{23}$  gas particles the futility of this endeavor becomes obvious rather quickly.

Instead, in statistical mechanics the state of the macroscopic system is described by macroscopic variables such as the number of particles  $N$ , the temperature  $T$  or the pressure  $P$ . They define the so called macrostate of the system. The corresponding microstate of the system is defined by the positions and momenta of all the particles in the system. The microstate can be represented as a point  $\mathbf{x}$  in the  $6N$  dimensional space of these positions and momenta. This space is called the phase space  $\Gamma$ . When the system evolves in time, the point  $\mathbf{x}$  moves through phase space along a trajectory. This time evolution is governed by deterministic laws, like Newtons equations of motion.

When determining any macroscopic observable  $A$  experimentally, one essentially takes the time average  $\langle A \rangle_t$  over a long time interval. This observable could be the energy of the system for example. The Law of large numbers tells us that for sufficiently large observation times the

average will converge to its true value [122]:

$$\langle A \rangle_t = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(t) dt. \quad (3.1)$$

This situation is mimicked by molecular dynamics (MD) simulations. They solve Newtons equations of motion to incrementally develop the system and calculate time averages of observables along the way. For discrete time steps, as is the case in MD simulations, one can rewrite eq. (3.1) as

$$\langle A \rangle_t = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N A(t_i), \quad (3.2)$$

where  $t_i$  are the discrete time steps. However, this is not the way we think about a system's thermodynamic properties in statistical mechanics, since in 1902 Gibbs suggested replacing the time average by an ensemble average [123]. An ensemble is a collection of points  $\mathbf{x}_i$  in phase space, where each point represents a possible microstate the system could be in, given a specific macrostate. The  $\mathbf{x}_i$  are distributed according to a probability distribution  $p(\mathbf{x}_i)$ . This  $p(\mathbf{x}_i)$  is determined by the set of fixed macroscopic parameters that one chooses (e.g.  $(N, V, E)$  or  $(N, V, T)$ ). The average of an observable then becomes the average over all the points in the ensemble weighted by the probability distribution of the states  $p(\mathbf{x}_i)$ :

$$\langle A \rangle_{\text{ens}} = \langle A \rangle = \sum_i p(\mathbf{x}_i) A(\mathbf{x}_i). \quad (3.3)$$

Using this definition of the observable average, one can sample phase space by randomly drawing points from the probability distribution  $p(\mathbf{x}_i)$ , as opposed to letting the system evolve in time. This constitutes the basic idea behind Monte Carlo simulations. They do not solve Newtons equations of motion to let the system develop in time, but instead randomly select microstates of the system to calculate ensemble averages. The probability distribution  $p(\mathbf{x}_i)$  is determined by the ensemble we choose. The so called microcanonical ensemble is defined by fixing the number of particles  $N$ , the volume  $V$  and the energy  $E$ . In equilibrium all microstates are equally probable. Therefore, their probability distribution is

$$p_{NVE}(\mathbf{x}_i) = \begin{cases} 1/\Omega & \text{if } \mathcal{H}(\mathbf{x}_i) = E \\ 0 & \text{otherwise,} \end{cases} \quad (3.4)$$

where  $\mathcal{H}$  is the Hamiltonian and  $\Omega$  is the total number of microstates at fixed  $(N, V, E)$ . This  $\Omega$  is also called the microcanonical partition function. Another statistical ensemble would be the canonical ensemble. It is defined by fixing the number of particles  $N$ , the volume  $V$  and the temperature  $T$ . This corresponds to a system that is connected to a heat bath, keeping it a constant temperature. The probability distribution of the microstates in the canonical ensemble is given by the Boltzmann distribution:

$$p_{NVT}(\mathbf{x}_i) = \frac{1}{Z} e^{-\beta \mathcal{H}(\mathbf{x}_i)} \quad (3.5)$$

where  $\beta = 1/k_B T$  is the inverse temperature and  $Z$  is the canonical partition function.  $Z$  acts as a normalization factor that ensures that the probability distribution integrates to one, meaning the probability of the system being in any state is equal to one. It is thus given by

$$Z = \sum_i e^{-\beta \mathcal{H}(\mathbf{x}_i)}. \quad (3.6)$$

As in the microcanonical ensemble, the canonical partition function is the sum over all microstates weighted by their probabilities. However, in the canonical ensemble the states do not

have equal probabilities as is the case in the microcanonical ensemble. Instead, they are weighted by the Boltzmann factor  $e^{-\beta\mathcal{H}(\mathbf{x}_i)}$ .

An important quantity in statistical mechanics is the entropy  $S$ . It is a measure of the number of possible microstates the system can be in, given a specific macrostate. The entropy is defined as:

$$S := -k_B \sum_i p(\mathbf{x}_i) \ln p(\mathbf{x}_i). \quad (3.7)$$

Using the probability distributions for the different ensembles (eqs. (3.4) and (3.5)), one can show that in the microcanonical ensemble the entropy is given by the Boltzmann formula:

$$S = k_B \ln \Omega \quad (3.8)$$

and in the canonical ensemble the expression for the entropy becomes

$$S = \langle E \rangle / T + k_B \ln Z. \quad (3.9)$$

From this formula, one can derive an expression for the Helmholtz free energy  $F$ . This establishes a connection between  $F$  and the canonical partition function  $Z$ :

$$-k_B T \ln Z = \langle E \rangle - TS =: F. \quad (3.10)$$

The logarithm of the partition function of a statistical ensemble is generally how the thermodynamic potential of the ensemble is defined. It marks the connection between the statistical mechanics description of the system and the thermodynamic description. Thermodynamic potentials are functions of the macroscopic parameters of the system and describe the equilibrium state of the system. They are used to calculate the thermodynamic observables and thus knowledge of the thermodynamic potential of a system is equivalent to knowing the system's thermodynamic properties. As the Helmholtz free energy can be derived from  $Z$ , all the thermodynamic information about the system is also stored in  $Z$ . Knowing  $Z$  is the key to connecting the microscopic description of the system to its macroscopic thermodynamic properties.

## 3.2 Standard Monte Carlo methods

The MC method is a numerical method that uses random sampling to calculate integrals. It was developed in the 1940s by von Neumann, Ulam and Metropolis to simulate neutron diffusion [124, 125]. Their research was part of the Manhattan Project for the development of nuclear weapons at Los Alamos. As of today, the MC method is arguably amongst the most important and useful numerical method for studying numerous scientific problems. This holds especially true in physics and in particular in statistical mechanics. The MC method is especially useful when calculating high-dimensional integrals, where deterministic methods become impracticable. This section will give a brief overview of standard MC methods. Starting with the basic simple sampling method, touching on Markov Chain processes and including the importance sampling Metropolis method. Understanding the basic principles of standard MC methods is important for understanding the more advanced MC methods. Following the description of the standard method the remainder of this chapter will focus on a description of the flat-histogram MC method Stochastic Approximation Monte Carlo that is used in this work.

### 3.2.1 Simple sampling Monte Carlo

As mentioned above, the MC method excels at calculating integrals. Take a function  $f(x)$  that we want to integrate over the interval  $[a, b]$ :

$$I = \int_a^b f(x) dx. \quad (3.11)$$

By expanding the integral with a probability distribution  $p(x)$  one can rewrite the integral as

$$I = \int_a^b \frac{f(x)}{p(x)} p(x) dx = \left\langle \frac{f}{p} \right\rangle. \quad (3.12)$$

Comparison with the summation representation of the average in eq. (3.3) reveals that the integration has effectively become the average of  $f(x)/p(x)$  where the  $x$  are distributed according to the probability distribution  $p(x)$ . By sampling  $N$  points  $x_{i=(1,\dots,N)}$  from a uniform probability distribution  $p(x) = 1/(b-a)$  for  $x \in [a, b]$ , and evaluating the function at the sample points it is possible to approximate the integral:

$$I = \frac{1}{N} \sum_i \frac{f(x)}{p(x)} + \mathcal{O}(N^{-1/2}) \quad (3.13)$$

$$\Rightarrow I \approx \frac{(b-a)}{N} \sum_{i=1}^N f(x_i). \quad (3.14)$$

This approach is called simple sampling. The error  $\epsilon$  of the approximation scales with  $\epsilon \propto 1/\sqrt{N}$ . Note, that this error scaling is independent of the dimensionality of the integral. In contrast, the error of traditional integration schemes like the midpoint-rule or the trapezoidal rule is dependent on the dimensionality of the integral. Since for these schemes each space component has to be partitioned separately a  $d$ -dimensional integral will have an error scaling of  $\epsilon \propto N^{-\nu/d}$ , where  $\nu$  is the one-dimensional error scaling. Therein lies the power of the MC method. It is a rather convenient method for calculating high-dimensional integrals with error scaling independent of the dimensionality of the integral.

In statistical mechanics the calculation of high-dimensional integrals is a frequently occurring problem. An example is the calculation of the average in phase space. Suppose we want to calculate the statistical average of the macroscopic quantity  $A(\mathbf{x})$  in the canonical ensemble (see eq. (3.3)). Following the simple sampling approach, one would then generate microstates  $\mathbf{x}_i$  with a uniform probability distribution. These microstates can for example be positions of beads in a model representation of a polymer, like the PRIME20 model. We then calculate the average  $\langle A(\mathbf{x}_i) \rangle_{NVT}$  over our sample. However, the specific shape of the probability distribution in the canonical ensemble creates an interesting problem. The Boltzmann distribution (eq. (3.5)) is sharply peaked at states  $\mathbf{x}_i$  that are in an energy minimum, making these states contribute the most to the average, while other states contribute almost nothing. Looking at systems investigated in this work, the system has 2 polymer chains with 20 residues each. A representation in the PRIME20 model results in 160 beads, which in turn leads to 480 coordinates in 3D space. In large systems that have this many degrees of freedom and where the number of possible microstates is huge, most of the generated states will have negligible contribution to the average as the Boltzmann factor will be near Zero for most of them. A far more efficient approach would be to generate microstates with a non-uniform probability distribution that is proportional to the Boltzmann distribution. This way, states that contribute the most to the average will be generated more often than those with a lower contribution. Doing so would greatly reduce the time needed for the integral approximation to converge to its true value. This idea is called Importance Sampling MC.

### 3.2.2 Importance sampling Monte Carlo

Importance sampling Monte Carlo aims to generate microstates with a probability distribution proportional to the probability distribution of the states in the investigated ensemble  $p_{\text{ens}}$ . In the case of the canonical ensemble the problem arises that  $p_{\text{ens}} = p_{NVT}$  includes the partition function  $Z$  (see eq. (3.5)) which is *a priori* unknown. Therefore,  $p_{NVT}$  cannot be calculated directly. In 1953 Metropolis et al. introduced the Metropolis algorithm which solves this problem

[126]. The Metropolis algorithm is a Markov Chain Monte Carlo (MCMC) method. A Markov Chain process is a sequence of states  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  in which the probability of transitioning to the state  $\mathbf{x}_i$  in the sequence only depends on the previous state  $\mathbf{x}_{i-1}$ . The idea of MCMC is derived from the Master equation. It describes the time evolution of the probability distribution between state  $p_{t-1}(\mathbf{x}_i)$  and  $p_t(\mathbf{x}_i)$  in a simulation and is given by:

$$p_t(\mathbf{x}_i) - p_{t-1}(\mathbf{x}_i) = \sum_{j \neq i} [p_{t-1}(\mathbf{x}_i)w(\mathbf{x}_j|\mathbf{x}_i) - p_{t-1}(\mathbf{x}_j)w(\mathbf{x}_i|\mathbf{x}_j)]. \quad (3.15)$$

Here  $w(\mathbf{x}_i|\mathbf{x}_j)$  is the conditional probability of reaching state  $\mathbf{x}_i$  when starting from state  $\mathbf{x}_j$ . As eq. (3.15) sums up the probability flow in and out of state  $\mathbf{x}_i$  it effectively describes conservation of probability in Markov Chain processes. Note, that in thermodynamic equilibrium the probability distribution  $p(\mathbf{x})$  is constant in time, which simplifies the Master equation to:

$$\sum_{j \neq i} [p(\mathbf{x}_i)w(\mathbf{x}_j|\mathbf{x}_i)] = \sum_{j \neq i} [p(\mathbf{x}_j)w(\mathbf{x}_i|\mathbf{x}_j)]. \quad (3.16)$$

The detailed balance condition is a specific realization and a more strict case of conservation of probability. It is valid for systems in equilibrium and requires that the probability of transitioning from state  $\mathbf{x}_i$  to state  $\mathbf{x}_j$  is equal to the probability of transitioning from state  $\mathbf{x}_j$  to state  $\mathbf{x}_i$ . Therefore, the detailed balance condition effectively gets rid of the sums in eq. (3.15). It is given by:

$$\frac{p(\mathbf{x}_i)}{p(\mathbf{x}_j)} = \frac{w(\mathbf{x}_i|\mathbf{x}_j)}{w(\mathbf{x}_j|\mathbf{x}_i)} \quad \forall i \neq j, \quad (3.17)$$

The transition probability  $w(\mathbf{x}_j|\mathbf{x}_i)$  in eqs. (3.15) to (3.17) is divided into two steps: the probability  $\alpha(\mathbf{x}_j|\mathbf{x}_i)$  of proposing the new state  $\mathbf{x}_j$  given the current state  $\mathbf{x}_i$  and the probability  $\tau(\mathbf{x}_j|\mathbf{x}_i)$  of accepting the proposed state. It follows, that  $w(\mathbf{x}_j|\mathbf{x}_i) = \alpha(\mathbf{x}_j|\mathbf{x}_i)\tau(\mathbf{x}_j|\mathbf{x}_i)$ . The mechanism of proposing the new state  $\mathbf{x}_j$  is specific to the system that is investigated. It will be discussed in detail for polymer systems in the PRIME20 model in section 3.3.3. It usually involves small variations of the current state  $\mathbf{x}_i$  and has to obey the microscopic reversibility principle:  $\alpha(\mathbf{x}_j|\mathbf{x}_i) = \alpha(\mathbf{x}_i|\mathbf{x}_j)$ . Thus, the proposition probabilities  $\alpha(\mathbf{x}_j|\mathbf{x}_i)$  cancel out of eq. (3.17) and the transition probability  $w(\mathbf{x}_j|\mathbf{x}_i)$  is given by the acceptance probability  $\tau(\mathbf{x}_j|\mathbf{x}_i)$ . Any choice of  $w(\mathbf{x}_j|\mathbf{x}_i)$  that satisfies the detailed balance condition (eq. (3.17)) will also satisfy the Master equation (eq. (3.15)). In an ergodic system, this  $w(\mathbf{x}_j|\mathbf{x}_i)$  will lead to a Markov Chain process that converges to the equilibrium distribution  $p(\mathbf{x})$  [127]. Ergodicity means, that any state in the system can be reached in a finite number of steps, utilizing a Markov Chain process in this case. Furthermore, if  $p(\mathbf{x})$  is the Boltzmann distribution  $p_{NVT}(\mathbf{x})$  (eq. (3.5)) then the partition function cancels out of the quotient on the left of eq. (3.17):

$$\frac{p_{NVT}(\mathbf{x}_i)}{p_{NVT}(\mathbf{x}_j)} = \frac{e^{-\beta\mathcal{H}(\mathbf{x}_i)}}{e^{-\beta\mathcal{H}(\mathbf{x}_j)}} = e^{-\beta\Delta\mathcal{H}}, \quad (3.18)$$

where  $\Delta\mathcal{H} = \mathcal{H}(\mathbf{x}_j) - \mathcal{H}(\mathbf{x}_i)$  is the energy difference between the states  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . This observation motivates the acceptance probability that was originally proposed by Metropolis et al.:

$$w(\mathbf{x}_j|\mathbf{x}_i) = \min[1, \exp(-\beta\Delta\mathcal{H})]. \quad (3.19)$$

It enables importance sampling without *a priori* knowledge of  $Z$  and only requires a calculation of the energy difference  $\Delta\mathcal{H}$  between the old and the new state.

Unfortunately, there are some downsides to the Metropolis algorithm. First,  $Z$  cancels out of the equations, which is the core idea of Metropolis MC. However, knowledge of the partition function would enable us to derive the entire canonical thermodynamics of the system (see eq. (3.10)). Thus, obtaining  $Z$  would be highly desirable. Another problem occurs in complex

systems with rough energy landscapes. MCMC methods are local search algorithms, which only make relatively small changes to the state of the system. In complex systems with multiple energy minima, this can lead to the algorithm getting stuck in low energy states that have a high probability density  $p(\mathbf{x})$ , because the acceptance probability  $w(\mathbf{x}_j|\mathbf{x}_i)$  penalizes transitions to states with higher energy (see eq. (3.19)). This can especially be a problem in systems with many degrees of freedom, like polymers in the PRIME20 model. In such systems, performing large steps in configuration space is difficult and thus, the ability to escape local energy traps is limited. If the simulation gets stuck, thermodynamic averages can not be calculated correctly, as the microstates will not be sufficiently sampled. Multiple advanced MC techniques try to address these shortcomings. Most notably, flat-histogram MC methods, like Umbrella Sampling [128], Multi-canonical MC [129, 130], Wang-Landau MC [131–136] and Stochastic Approximation MC [45, 56, 131, 135–140].

### 3.3 Stochastic Approximation Monte Carlo

SAMC was developed by Liang et al. [45] in 2007. It is very similar to the Wang-Landau Monte Carlo (WLMC) method and can be seen as its mathematical pendant. Both of these methods are part of the flat-histogram MC methods. As such, they aim to sample the energy space of a system in a uniform manner. By doing so, they achieve an even visitation histogram of the energy states in the system. The idea behind SAMC and WLMC is to approximate the microcanonical density of states (mDOS)  $g(E)$ , which counts the number of states with energy  $E$ .<sup>(3)</sup> A major motivation behind obtaining the mDOS is that with it one can calculate the partition function  $Z$  by replacing the sum over states (see eq. (3.6)) with a sum over energies:

$$Z = \sum_E g(E) e^{-\beta E}. \quad (3.20)$$

Since  $g(E)$  is independent of the temperature it can be used to find all thermodynamic properties of the system at any temperature. From observable averages  $A(E)$  collected in the microcanonical ensemble during the simulation runs it is possible to calculate the canonical ensemble averages as

$$\langle A \rangle(T) = \frac{1}{Z} \sum_E g(E) e^{-\beta E} A(E). \quad (3.21)$$

It is worth noting, that the mDOS can depend on multiple parameters, not only the energy  $E$ . It can be used for even sampling of any parameter space and their combinations [140–143]. However, in the work at hand only energy dependence was considered and the mDOS will be denoted as  $g(E)$ . Furthermore, configurational energy space was sampled ( $E = U$ ) because kinetic information is not directly accessible in Monte Carlo simulations. The ”missing” kinetic part of the mDOS will be included post-simulation. How this is done and its effects on derived thermodynamic variables will be discussed in more detail in section 3.4.

In order to motivate the algorithm of flat-histogram MC approaches, let us consider a random walk through configuration space  $\Gamma(\mathbf{x})$  of the system. The probability of arriving at a state  $\mathbf{x}'$  which has an energy  $E'$  in the interval  $[E', E' + \Delta E)$  is proportional to the number of states  $g(E')\Delta E$  in that energy interval. Vice versa, the probability to start from a specific state  $\mathbf{x}$  with energy  $E$  is proportional to  $1/(g(E)\Delta E)$ . Thus, performing MCMC with an acceptance probability that is proportional to the ratio of the density of states of the old and the new state  $g(E)/g(E')$  will lead to uniform sampling of energy states in the system. The problem at hand now is that one does not know  $g(E)$ . SAMC aims to approximate  $g(E)$  in an iterative manner, which will be described in the following section.

---

<sup>(3)</sup>This density of states  $g(E)$  is the same as the density of states  $\Omega$ , that was introduced in section 3.1.1. The change to  $g(E)$  is made at this point to keep nomenclature consistent with literature [45]

### 3.3.1 Algorithm of Stochastic Approximation Monte Carlo

We start by assuming a microscopic probability density  $\Psi(\mathbf{x})$  on configuration space  $\Gamma(\mathbf{x})$ . Furthermore, we chose an energy interval  $[E_{\min}, E_{\max}]$  with  $M$  discrete energy states  $E_i$ . The chosen minimum energy  $E_{\min}$  can be below the ground state energy  $E_g$  of the system. The energy states can either be intrinsic to the system or, for systems with continuous energy range and states, constructed by appropriate binning of adjacent energies [144]. This set of energies leads to a unique partitioning of  $\Gamma$  [130]:

$$g(E_i) = \int_{E_i \leq U(\mathbf{x}) < E_{i+1}} \Psi(\mathbf{x}) d\mathbf{x}. \quad (3.22)$$

Since SAMC aims to approximate  $g(E_i) = g_i$ , there is a current working estimate of the mDOS at simulation time step  $t$  which will be denoted  $\hat{g}_i(t)$ . We start with an unbiased guess for  $\hat{g}_i$ , for example  $\hat{g}_i(t=0) = \hat{g}_{i,0} = 1, \forall i$  [45]. However, it has been shown, that for fast convergence it is beneficial to have a good initial guess of  $\hat{g}_{i,0}$ . A good first estimate would be the mDOS of previous simulation runs of similar systems. Alternatively, one can utilize short preliminary runs to get a rough approximation of  $\hat{g}_i$  [140]. Furthermore, it is advantageous to work with the logarithm  $\ln g_i$ , because  $g_i$  varies over multiple orders of magnitude. Next to the estimate  $\ln \hat{g}_i(t)$ , the simulation keeps track of the histogram  $H_i(t)$  of visited energy states. It counts the number of times the system visited an energy state in the interval  $[E_i, E_{i+1})$  within  $t$  simulation time steps. At the beginning of the simulation we set  $H_i(t=0) = 0, \forall i$ . The simulation then performs MCMC steps until a finalization criterion is met. Usually, this criterion is a maximum number of simulation time steps  $t_{\max}$ . Let's assume the simulation at time step  $t$  proposes a move from state  $\mathbf{x}_a$  to  $\mathbf{x}_b$ , which have the energies  $E_a$  and  $E_b$  respectively. This move will be accepted with the probability:

$$w(\mathbf{x}_a, \mathbf{x}_b) = \min \left\{ 1, \frac{\hat{g}_a(t)}{\hat{g}_b(t)} \right\}. \quad (3.23)$$

After accepting or rejecting the move, the new microstate of the system is  $\mathbf{x}^*$  with the energy  $E^*$ . The simulation then updates the mDOS estimate  $\ln \hat{g}_i(t)$  and the histogram  $H_i(t)$  at the position of  $E^*$  according to:

$$\ln \hat{g}_i(t+1) \rightarrow \ln \hat{g}_i(t) + \gamma_t \delta(E^*, E_i), \quad (3.24)$$

$$H_i(t+1) \rightarrow H_i(t) + \delta(E^*, E_i). \quad (3.25)$$

Here  $\delta(E^*, E_i)$  is the Kronecker delta function that is equal to one if the energy of the new state  $E^*$  is in the energy interval  $[E_i, E_{i+1})$  and zero otherwise. The modification factor  $\gamma_t$  decreases over the course of the simulation according to

$$\gamma_t = \min \left( \gamma_0, \frac{t_0}{t} \right). \quad (3.26)$$

The convergence of the SAMC algorithm has been proven mathematically by Liang et al. [45]. For convergence  $\gamma_t$  has to satisfy

$$\sum_{t=1}^{\infty} \gamma_t = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \gamma_t^{\zeta} < \infty, \quad (3.27)$$

for some  $\zeta \in (1, 2)$ . In the convergence criteria for the time development of  $\gamma_t$  lies the main difference between SAMC and WLMC. While in WLMC the algorithm waits for the histogram to be sufficiently flat before reducing  $\gamma_t$ , SAMC reduces  $\gamma_t$  at a predetermined rate. Though it might be more intuitive to wait for convergence, it has been shown that SAMC leaves a non-zero

residual error when  $\gamma_t$  is changed exponentially [145]. Furthermore, in WLMC the time to reach a desired precision is unpredictable, whereas in SAMC the simulation reaches the minimum value of  $\gamma_t$  in a predefined number of moves. Another advantage of SAMC over WLMC is that the energy range over which  $g(E)$  is to be approximated does not have to be fully within the boundaries of the energy range of the investigated system but can in fact exceed it. This comes in handy if the ground state energy of the system is unknown.

Finally, it should be noted, that in the SAMC scheme  $\ln g(E)$  converges to the true value of  $\ln g(E)$  only up to an unknown constant  $S_0$  [130]. That is, because the acceptance probability in eq. (3.23) depends on the ratio of the  $g(E)$  of different states.

### 3.3.2 Implementation details and simulation parameters

The SAMC simulation code used in this work was written in the C++ programming language. It is based on the simulation code by Arne Böker [53, 61], which he developed in the C programming language for the simulation of single chain homopolymers in the PRIME20 model. The change in programming language (C  $\rightarrow$  C++) is due to author preference. The simulation code that is used in this work implements the option to simulate multi-chain systems into the previous program version. This change demanded adjustments to a number of features in the base code. The most important adjustments are discussed below.

#### Change to multi-chain systems

First, there is the issue of keeping the density in the system constant. On the one hand this is important for keeping the system at physiological conditions. On the other hand this assures that the polymer chains are kept in close proximity to each other which lets them interact frequently. If the system is left unconfined, initially close chains would most probably end up far apart from each other and outside their interaction range after a few MC steps. To keep the chains close to each other, they are put inside a simulation box with periodic boundary conditions (PBC). This means that the simulation box is replicated throughout space forming an infinite lattice. As particles move through the simulation box, their periodic replicas in the other boxes move in the exact same way. If a particle leaves its box, a respective replica enters the box from the opposite side. At all times, one replica of each particle will be in the primary simulation box. Thus, it is only necessary to store coordinates of the particles in the primary simulation box, not those of all the periodic replicas in the image boxes. By using PBC, the system is effectively infinite and not influenced by boundary effects, as would be the case if the simulation box had solid border walls.

In systems with PBC, extra attention has to be paid when calculating distances between particles. Multiple versions of the same particle exist, i.e. one in each replica box. Hence, only the shortest distance to a particle, which means to the closest version of this particle, should be considered in distance calculations. This is called the minimum image convention. With this in mind, the distance between two particles is calculated in the following way:

$$d_{\text{PBC}}^{(\alpha)} = d_{\text{SB}}^{(\alpha)} - L \operatorname{sgn}(d_{\text{SB}}^{(\alpha)}) \left\lfloor \left| \frac{d_{\text{SB}}^{(\alpha)}}{L} + 0.5 \right| \right\rfloor, \quad \alpha \in (1, \dots, D). \quad (3.28)$$

Here  $d_{\text{PBC}}^{(\alpha)}$  is the shortest distance between the two particles when taking PBC into account and  $d_{\text{SB}}^{(\alpha)}$  is the distance calculated in the primary simulation box. The superscript  $(\alpha)$  indicates separate calculations for each coordinate in a Cartesian coordinate system with Dimension  $D$ .  $L$  is the side length of the simulation box and the operator  $\lfloor x \rfloor$  is the floor function that returns the largest integer smaller than  $x$ .

Distance calculations have to be performed multiple times per simulation time step. For example, it is required in the calculation of the acceptance probability of a proposed MC move.

For this one needs to determine the energy of the old and the new state (see eq. (3.23)). Energy in the PRIME20 model depends on the distance between particles (see section 2.3). A brute-force approach would be to calculate all distances between all particles in the system and then check which ones are within each other's interaction range. This would require  $O(N^2)$  operations, where  $N$  is the number of particles in the system. This is computationally expensive and unnecessary, because most particles are too far apart from each other to contribute to the energy of the system. To reduce the number of distance calculations, the simulation code uses a cell structure and a linked list [60, 146, 147]. A linked list is a type of neighbor list that acts as a bookkeeping device to keep track of neighboring particles. The idea behind the method is to divide the simulation box into square cells which have a side length  $L_{\text{cell}}$  that is greater than the largest interaction distance  $d_{\text{huge}}$  in the system. In the case of PRIME20,  $d_{\text{huge}} = 7.4 \text{ \AA}$  (see table A.4). By choosing  $L_{\text{cell}} > d_{\text{huge}}$ , particles that are within each other's interaction range will be in the same cell or in neighboring cells. During the course of the simulation, a list is maintained that keeps track of which particle belongs to which cell. In this way, when calculating distances between particles, one only needs to go through the lists of particles in the relevant cells in order to get all particles relevant for the distance calculation. This massively reduces the computational effort of distance calculations to  $O(N)$  operations [60]. Consider the 2D example in fig. 3.1(a), where the simulation box is divided into  $5 \times 5$  cells. If the particle of which the energy is to be calculated is located in cell 7 then only particles in cells (1,2,3,6,8,11,12,13) have to be considered. Extra attention has to be paid when looking for neighbors of a particle close to the border of the simulation box. Their neighbors can be located in cells on the opposite side of the box, due to the PBC. Looking again at the example in fig. 3.1(a), neighboring cells to cell 11 include cells 10,15 and 20.

The linked cell list method has a sophisticated way of listing the particles in the cells which works as follows: First, one sorts every particle into its appropriate cell. As this step has a low computing cost, it can be performed after every position update. During the sorting process two lists are created: a Head-Of-Chain list (HOCL) and a Linked-List (LL). The HOCL contains one entry for each cell. Each entry contains the largest particle index  $q_1$  in the respective cell. This marks the entry position  $\text{LL}[q_1]$  into the linked list.  $\text{LL}[q_1]$  then contains the index  $q_2$  of the particle in the cell with the next-lower index, which in turn contains the next-lowest index  $q_3$ , and so on. The linked list terminates when reaching an element with index Zero. This indicates that there are no further particles in the cell. As an example, again consider the system with  $5 \times 5$  cells from before (fig. 3.1). Let's assume, that 12 particles are distributed between cells 7 and 8 as seen in fig. 3.1(b). Cell 7 contains particles (2,5,6,9,10,11) and cell 8 contains particles (1,3,4,7,8,12). This results in the two lists, HOCL and LL, being set up as shown in table 3.1. When looking for the particles in cell 7, one would start the LL at position  $\text{HOCL}[7] = 11$ . Following the linked list through the cells, one would find the indices of all particles in cell 7 by following the linked list:

$$\text{LL}[11] = 10 \rightarrow \text{LL}[10] = 9 \rightarrow \text{LL}[9] = 6 \rightarrow \text{LL}[6] = 5 \rightarrow \text{LL}[5] = 2 \rightarrow \text{LL}[2] = 0.$$

In the same way one can retrieve the indices of all particles in cell 8 by starting in the LL at position  $\text{HOCL}[8] = 12$  and going through the linked list:

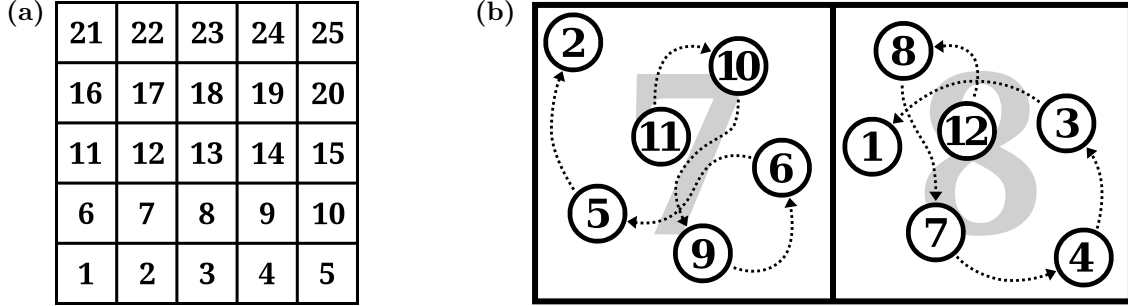
$$\text{LL}[12] = 8 \rightarrow \text{LL}[8] = 7 \rightarrow \text{LL}[7] = 4 \rightarrow \text{LL}[4] = 3 \rightarrow \text{LL}[3] = 1 \rightarrow \text{LL}[1] = 0.$$

This way, the linked cell list method allows for efficient distance calculations in systems with a finite simulation box size, including systems with PBC.

Another change made to the simulation code due to the accommodation of multi-chain systems was the implementation of additional MC moves. Since a different relative distances and orientations of the chains constitutes a different microstate of the system, it is necessary to introduce moves that change the position and orientation of entire chains. The implementation of all MC moves utilized in the simulation will be discussed in more detail in section 3.3.3.

**Table 3.1:** Head-of-chain array ("HOCL") and Linked List ("LL"). For cell  $i$  starting at  $LL[q_1]$  with  $q_1$  specified in  $HOCL[i]$  one can follow the trail through LL where each element specifies the index of the next particle in the cell.

Index	1	2	3	4	5	6	7	8	9	10	11	12
HOCL							11	12				
LL	0	0	1	3	2	5	4	7	6	9	10	8



**Figure 3.1:** (a) Example of a 2D simulation box divided into 25 cells. (b) The linked List in table 3.1 is used to find the particles in cells 7 and 8.

### Simulation parameters

Interaction potentials in the PRIME20 model are discontinuous (see section 2.3). This leads to the energy of a microstate only assuming discrete values. Therefore, the energy scale was divided into bins of width  $d_{\text{bin}} = 0.1$ . In homopolymer systems of PolyQ chains, the energy contribution of a single side chain contact is  $\varepsilon_{\text{SC}} = -0.08$ . The energy contribution of a HB is  $\varepsilon_{\text{HB}} = -1.00$ . From this it follows, that the energy  $E$  of a microstate can assume values that are multiples of 0.08, when  $E \in [-1; 0]$ . Since  $0.08 \cdot 13 = 1.04$ ,  $E$  can reach values that are multiples of 0.04, when  $E < -1$ . However, especially at shorter chain lengths, these discrete energy values can not be fully populated. Thus,  $d_{\text{bin}} = 0.1$  was chosen, as it yields a good compromise between resolution and population of the energy states. The division of energy space with  $d_{\text{bin}} = 0.1$  was kept the same for the investigation of the heteropolymer systems of A $\beta$  and PTH. It should be mentioned that this energy binning is rather basic and does not account for the fact that it groups microstates with different microscopic probability densities into the same energy bin. Because the sampling takes place in the microcanonical ensemble, it is assumed that the probabilities of the microstates in the same energy bin are equal. But especially at low energies, microstates with an energy at the lower border of the energy bin can have a much lower probability of occurring than a state with an energy at the upper border of the energy bin. A way to alleviate this problem would be to interpolate  $\ln \hat{g}_i$  between the energy states  $E_i$  and  $E_{i+1}$ . This would allow for a more accurate accumulation of  $\ln \hat{g}_i$  by assigning different mDOS to different states in the same energy bin when calculating the acceptance probability of the MC steps [148]. However, this requires a more complex implementation of the acceptance probability calculation and was not done in this work.

As described above in section 3.3.1, the choice of  $\gamma_0$  and  $t_0$  is important for the convergence of the SAMC algorithm. The initial modification factor  $\gamma_0$  was chosen in the interval  $[0.001, 0.004]$ . Convergence of simulations of larger polymer systems benefited from larger values of  $\gamma_0$  as the mDOS has to be accumulated over a large energy range. On the other hand, when improving upon the accuracy of previous simulation runs of which the  $\hat{g}_i$  was used as initial guess for the new simulation run, a smaller  $\gamma_0$  was chosen to reach a smaller target  $\gamma_t$ . As was shown by Shakirov, the accuracy of the simulation results also depends on the ratio of  $t_0$  to the number of energy bins  $N_{\text{bin}}$  in the sampling range [140]. To achieve high accuracy,  $t_0$  was chosen according

**Table 3.2:** Simulation box side lengths  $L$  for the different polymer systems.  $N_{\text{res}}$  is the number of residues in the polymer chain, also referred to as the chain length.

Polymer	$N_{\text{res}}$	$L$
Polyglutamine	14,16,18,20,22,24,26	112.5 Å
	28,36	150.0 Å
Amyloid- $\beta$	40,42	300.0 Å
PTH	34,42,84	300.0 Å

to the formula:

$$t_0 = \frac{N_{\text{bin}}}{N_{\text{bead}}} \nu, \quad \nu \in (1.5, 5.0). \quad (3.29)$$

Here  $N_{\text{bead}}$  is the number of beads in the system. Its inclusion in the calculation of  $t_0$  accounts for the fact that one MC step, after which  $\gamma_t$  is recalculated, involves  $N_{\text{bead}}$  MC moves. This ensures that every bead in the system has a chance of being moved during one MC time step. The choice of the factor  $\nu$  again depends on the accuracy of previous results that were used as initial guess for the new simulation run. A smaller  $\nu$  leads to a smaller  $t_0$ , which in turn leads to a smaller target  $\gamma_t$  and thus a potentially higher accuracy of the simulation results. This comes with the trade-off, of accumulating less mDOS over the course of the simulation, which leaves less room for adjusting errors in the initial guess of  $\hat{g}_i$ .

The three different polymer systems (PolyQ, A $\beta$  and PTH) have different amino acid sequences and different chain lengths (number of residues per chain)  $N_{\text{res}}$ . The sequences are:

$$\text{PolyQ}_{N_{\text{res}}} = (\text{Q}) \cdot N_{\text{res}} \quad (3.30)$$

$$\text{A}\beta(1 - 42) = \text{DAEFR}_5 \text{HDSGY}_{10} \text{EVHHQ}_{15} \text{KLVFF}_{20} \text{AEDVG}_{25} \text{SNKGA}_{30} - \text{IIGLM}_{35} \text{VGGVV}_{40} \text{IA} \quad (3.31)$$

$$\text{PTH}(1 - 84) = \text{SVSEI}_5 \text{QLMHN}_{10} \text{LGKHL}_{15} \text{NSMER}_{20} \text{VEWLR}_{25} \text{KKLQD}_{30} - \text{VHNFV}_{35} \text{ALGAP}_{40} \text{LAPRD}_{45} \text{AGSQR}_{50} \text{PRKKE}_{55} \text{DNVLV}_{60} - \text{ESHEK}_{65} \text{SLGEA}_{70} \text{DKADV}_{75} \text{NVLTk}_{80} \text{AKSQ}. \quad (3.32)$$

Subsets and variants of these proteins are identified by the positions of the start and end residues. For example A $\beta$ (1 – 40) is the first 40 residues of the full A $\beta$  sequence. The colors in eqs. (3.30) to (3.32) help differentiate between the different chain lengths that were used in the simulations. The simulation box side lengths  $L$  were chosen in such a way that the systems were at a molecular concentration of  $c \approx 1$  mM. This was done to ensure that the concentration in the systems were close to *in vitro* experiment conditions. The systems that were simulated with the chosen  $N_{\text{res}}$  and  $L$  are shown in table 3.2.  $L$  has not the same value for the different simulated systems, because the different lengths of the polymer chains put different restrictions on the minimal box length  $L_{\text{min}}$ . If  $L_{\text{min}}$  is too small, a single chain could stretch from one side of the simulation box to the other and interact with itself. This in turn would lead to an overestimation of the number of intra-chain contacts and other undesirable effects.

Simulations were performed on the high performance computing (HPC) cluster of the theoretical physics faculty of the Martin-Luther University Halle-Wittenberg. On Intel Xeon E5-2640 CPUs with 2.40 to 2.60 GHz and a minimum of 16 GB of RAM, simulations runs took approximately 60 days on average to converge on a mDOS with a precision of  $\gamma_t < 10^{-7}$ . 8 or 16 simulation runs were performed in parallel on the cluster. The number of parallel runs was chosen according to the number of available cores on the cluster. Converged simulation runs over the same energy window were averaged. Multiple energy windows were combined to form the total mDOS of the system.

### Additional software tools

Analysis of the simulation results was performed with the help of various software tools. The most important of which are listed here:

Data analysis was mainly performed in the python programming language using the NumPy library [149]. Additional analysis, was performed in C++ [150]. GNU bash scripts were used for automation of data handling and minor data analysis tasks [151]. Plots were created using the Matplotlib library in python. Other diagrams and figures were created and edited using GIMP (GNU Image Manipulation Program) [152] and Inkscape [153]. Polymer structures were visualized with PyMOL [154].

### 3.3.3 Monte Carlo moves

The SAMC algorithm is a type of MCMC method. As such, it requires a set of MC moves to change the current state of the system. A thorough description of different MC moves for polymer systems can be found in [155]. The following MC moves were implemented in the simulation code: local displacement of a single bead, pivot rotation moves and whole chain moves.

The most commonly selected MC move is the local displacement of a single bead. It is selected 50 times as often as the other MC moves. The move is a small local change of the position of a single randomly selected bead in a random direction. The maximum distance of displacement in each cardinal direction is set to  $0.02 \text{ \AA}$ . It is the only move type to change local distances and bond lengths between beads. Thus, it is an essential move in order to fully sample configuration space. Furthermore, the potentials in the PRIME20 model are short-ranged and the energy of the system is highly dependent on these local distances.

The Pivot rotation move changes the configuration of the polymer chain by rotating the  $\Psi$  or  $\Phi$  angle of a randomly selected residue by a random amount (the angles are shown in fig. 2.4(a)). The maximum angle of rotation is set to  $\pi/3$ , which optimizes the move's acceptance rate while still potentially producing large changes to the conformation with large energy jumps. In single chain systems, the choice of the chain end that is rotated is arbitrary, since it does not make a difference in the microstate of the system. However, in multi-chain systems, rotating different sides of the chain leads to different microstates due to the relative positions of beads of different chains. Thus, in the multi-chain systems the chain ends on both sides of the selected angle can potentially be selected for rotation. The pivot moves do not change any bond lengths. Instead, they are used to make large scale changes to the configuration of the polymer chain. This leads to a lower acceptance rate of the pivot move compared to the local displacement move. This is especially noticeable when the system is in a low energy state, where the steric restrictions are high and the free volume is low.

The third kind of MC moves are the whole chain moves. They consist of two types: the whole chain translation and the whole chain rotation. The translation move is a large scale random displacement of the entire chain in a random direction. In each cardinal direction, a random distance in the interval  $[-d_{\max}, d_{\max}]$  is selected, where  $d_{\max}$  is set to around  $0.5L$ . The whole-chain rotation move is a rotation of the entire chain around its geometric center and around a randomly selected rotation axis. The maximum angle of rotation is  $2\pi$ . These whole chain moves are important to change the relative position and orientation of the chains in multi-chain systems.

Acceptance rates varied for all move types over the energy range of the system. In high energy regions, the acceptance rates were high, because there are less steric restrictions due to the hard sphere repulsion. In low energy regions, the acceptance rates were low, because there are more steric restrictions and less free volume. This holds especially true for the large scale moves.

### 3.4 Observables

Simulations were performed in two stages: the SAMC runs and the production runs. The SAMC runs were used to estimate the mDOS of the system. The production runs then used the approximated  $\hat{g}_i$  to uniformly sample the energy space of the system. Production runs did not change the mDOS any further. Instead, they were used to calculate the observables of the system in dependence of the energy. As mentioned before, in MC only the potential energy of the system is accessible. This section will discuss how to transition from the configurational microcanonical ensemble to the full-energy microcanonical ensemble. Furthermore, conversion of observables to the canonical ensemble will be discussed. Finally, the observables that were calculated in the simulation runs will be presented, which are grouped into thermodynamic and structural observables.

#### 3.4.1 Transition between ensembles

During the SAMC runs, observables  $\langle A \rangle(U)$  were collected in the microcanonical ensemble. To convert the microcanonical ensemble averages to the canonical ensemble, one can use the mDOS  $g(U)$  to calculate the canonical ensemble averages according to eq. (3.21). However, the microcanonical ensemble is usually defined in phase space, not in configuration space. Thus, one needs to include the kinetic part of the mDOS to get from the configurational microcanonical ensemble ( $NVU$ ) to the full-energy microcanonical ensemble ( $NVE$ ). Here,  $E = U + K$  is the total energy of the system with the potential energy  $U$  and the kinetic energy  $K$ . Shakirov et al. presented a method of obtaining the mDOS in phase space from the mDOS in configuration space via the convolution of the two [142]:

$$\begin{aligned} g(E) &\propto \sum_U (E - U)^{f/2-1} g(U) \Theta(E - U) \\ &= \sum_U \exp \left[ \left( \frac{f}{2} - 1 \right) \ln[E - U] + \ln[g(U)] \right] \Theta(E - U). \end{aligned} \quad (3.33)$$

Here,  $\Theta(x)$  is the Heaviside step function and  $f$  is the number of degrees of freedom of the system. In the PRIME20 model, a system of two polymer chains, each with  $N_{\text{res}}$  residues, has  $f = 24 N_{\text{res}} - 3$  degrees of freedom. This is because each of the  $N_{\text{res}}$  residues in the 2 chains has 4 beads, each with 3 degrees of freedom. Three coordinates are restricted by conservation of momentum. As was the case for the configurational mDOS,  $g(E)$  is only known up to an unknown constant, hence the proportionality in eq. (3.33).

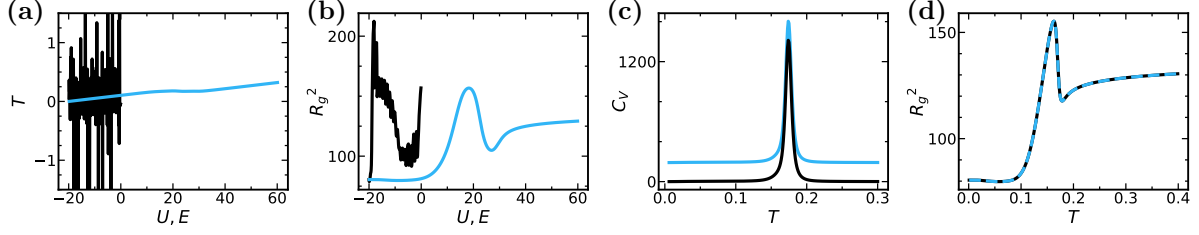
Observables in the configurational ensemble can be transformed to the full-microcanonical ensemble using the conditional probability  $p(E|U)$  of finding the system in a state with energy  $E$  given a potential energy  $U$ :

$$A(E) = \sum_U A(U) p(E|U) \quad (3.34)$$

$$\text{with } p(E|U) = \frac{(E - U)^{f/2-1} g(U) \Theta(E - U)}{\sum_U (E - U)^{f/2-1} g(U) \Theta(E - U)}. \quad (3.35)$$

Next to the microcanonical ensemble, the canonical ensemble is of interest. As already discussed in section 3.3, canonical ensemble averages can be obtained by calculating the partition function  $Z$  (eq. (3.20)). With knowledge of  $Z$  the canonical ensemble averages are calculated using the mDOS  $g(E)$  according to eq. (3.21). By replacing the total energy  $E$  with the potential energy  $U$  in both equations one gets configurational results in the canonical ensemble.

Thus, there are four different ensembles up for consideration: the configurational and full-energy microcanonical ensemble and their respective canonical ensembles. However, comparing



**Figure 3.2:** Comparison of thermodynamic observables in the configurational (black) and full-energy (blue) microcanonical ensemble and their respective canonical ensembles. (a) Temperature  $T$  as a function of energy  $E$ . (b) Squared radius of gyration  $R_g^2$  as a function of energy  $E$ . (c) Canonical heat capacity  $C_V$  as a function of temperature  $T$ . (d)  $R_g^2$  as a function of temperature  $T$ . Values of  $U$ ,  $E$ ,  $T$  and  $C_V$  are given in PRIME20 model units, while  $R_g^2$  is given in  $\text{\AA}^2$ .

the configurational and full-energy ensembles shows, that in the canonical ensemble the distinction between configurational and full-energy ensemble is not required. A comparison between the different ensembles is shown in fig. 3.2. The temperature (fig. 3.2(a)) and the squared radius of gyration are common observables in the microcanonical ensemble. The representation in the configurational microcanonical ensemble is colored in black, while the full-energy microcanonical ensemble is assigned the color blue. It is immediately evident, that both graphs look very different and hold different information. Therefore, both the configurational and the full-energy microcanonical ensemble will be used in the thermodynamic analysis of the simulation results. For the canonical ensemble, fig. 3.2(c) shows the canonical heat capacity and fig. 3.2(d) again shows the radius of gyration, but this time as a function of temperature. In the full-energy canonical ensemble, the heat capacity is shifted upwards by a factor of  $f/2$ , compared to the configurational canonical ensemble. This is due to the fact that the heat capacity is a measure of the fluctuations of the energy of the system and the expected kinetic energy gain is  $k_B T/2$  per degree of freedom. Apart from this shift, the shape of the two heat capacity curves is identical. Since in the analysis, the maxima positions of the heat capacity are of interest, the shift does not affect the results. For the radius of gyration, the change from the configurational to the full-energy canonical ensemble does not affect the results at all. The situation is similar for other structural observables. Thus, a distinction between the configurational and full-energy canonical ensemble is not necessary for the analysis of the simulation results.

### 3.4.2 Thermodynamic observables

The primary result of the SAMC simulation is the mDOS  $g(E)$ . Since  $g(E)$  spans hundreds of orders of magnitude, the simulation operates using its logarithm  $\ln g(E)$ , which is directly related and proportional to the system's entropy:

$$S(E) = k_B \ln g(E) + S_0. \quad (3.36)$$

Here,  $k_B$  is the Boltzmann constant and  $S_0$  is the unknown constant up to which SAMC approximates the mDOS. This  $S_0$  originates from the MC acceptance criterion (eq. (3.23)) which uses the ratio of the  $g(U)$  of different states. Entropy is a measure for the number of microstates that correspond to a certain energy. By setting  $k_B = 1$  the logarithm  $\ln g(E)$  directly represents the entropy  $S(E)$ .

In order to investigate the thermodynamic properties of the system one turns to its deriva-

tives, namely the temperature  $T(E)$  and the microcanonical heat capacity  $c(E)$ :

$$T(E) = \left( \frac{\partial S}{\partial E} \right)^{-1} \quad (3.37)$$

$$c(E) = \left( \frac{\partial T}{\partial E} \right)^{-1} = -\frac{1}{T^2} \left( \frac{\partial^2 S}{\partial E^2} \right)^{-1}. \quad (3.38)$$

The configurational versions of the entropy, temperature and heat capacity are calculated from the configurational mDOS  $g(U)$  in the same way as described above, by replacing  $E$  with  $U$  in the equations.

In the canonical ensemble, the average energy  $\langle E \rangle(T)$  and the heat capacity  $c_V(T)$  are of interest. The average energy is calculated as follows:

$$\langle E \rangle(T) = \frac{1}{Z} \sum_E E g(E) e^{-\beta E} = -\frac{\partial \ln Z}{\partial \beta}, \quad (3.39)$$

where  $\beta = 1/k_B T$  is the inverse temperature. The heat capacity is then calculated as the derivative of the average energy:

$$C_V(T) = \frac{\partial \langle E \rangle}{\partial T} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2} \quad (3.40)$$

With knowledge of the mDOS  $g(E)$ , the partition function  $Z$  can be calculated as shown in eq. (3.20). This allows any observable that was collected in the simulation's microcanonical ensemble to be transformed to the canonical ensemble using eq. (3.21). Consequently, the simulation results can be presented as functions of both the energy and the temperature of the system, offering a comprehensive perspective of its thermodynamic behavior. However, the temperature dependence of the observables is of particular interest, as it more closely relates to experimental conditions.

### Identifying phase transitions

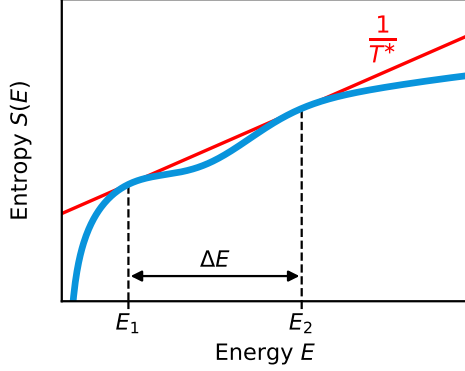
With the thermodynamic quantities described above, it is possible to identify phase transitions (PTs) in the system. According to the Ehrenfest classification, a PT is of  $n$ -th order if the thermodynamic potential exhibits a discontinuity in its  $n$ -th derivative with respect to some thermodynamic variable, while all lower-order derivatives remain continuous. In the canonical ensemble, the relevant thermodynamic potential is the Helmholtz free energy  $F(T, V, N)$ , which can be obtained from the logarithm of the partition function  $Z$  and can therefore be computed from the mDOS (see eqs. (3.10) and (3.20)). The first derivative of  $F$  with respect to  $T$  is proportional to the average energy  $\langle E \rangle(T)$ , as can be derived from eq. (3.39):

$$\langle E \rangle(T) = -\frac{\partial \ln Z}{\partial \beta} = F - T \frac{\partial F}{\partial T}, \quad (3.41)$$

where  $F = -k_B T \ln Z$  and  $\beta = 1/(k_B T)$ . The second derivative of  $F$  with respect to  $T$  is used to calculate the heat capacity:

$$C_V(T) = \frac{\partial}{\partial T} \left( F - T \frac{\partial F}{\partial T} \right) = -T \frac{\partial^2 F}{\partial T^2}. \quad (3.42)$$

In a first-order PT,  $\langle E \rangle$  is discontinuous at the PT temperature  $T^*$ , leading to a divergence of  $C_V(T)$  at  $T^*$ . In finite systems, however,  $C_V(T)$  does not diverge but instead exhibits a sharp peak. For a second-order PT,  $C_V(T)$  diverges in the thermodynamic limit, while in finite systems, this divergence manifests as a broad peak.



**Figure 3.3:** Schematic example of an entropy curve  $S(E)$  featuring a convex intruder.  $S(E)$  is colored in blue, while the tangent to the convex intruder is colored in red.

In addition to the canonical ensemble, the microcanonical ensemble can also be used to identify PTs. Here, the shape of the entropy  $S(E)$  is directly analyzed. The slope of  $S(E)$  is proportional to the inverse temperature  $1/T$ , as can be seen from eq. (3.37). For a given energy  $E_1$  and the corresponding entropy  $S(E_1)$ , the temperature  $T_1$  defines a tangent to  $S(E)$  with slope  $1/T_1 = \frac{dS}{dE}|_{E_1}$ . The equation of said tangent is given by:

$$\begin{aligned} S - S(E_1) &= \frac{E - E_1}{T_1} \\ \Rightarrow E - T_1 S &= E_1 - T_1 S(E_1) = F_1. \end{aligned} \quad (3.43)$$

Thus, all points on the tangent have the same free energy  $F_1$ . If  $S(E)$  exhibits a convex region, then there exists a tangent on  $S(E)$  with slope  $1/T^*$  and two contact points  $E_1$  and  $E_2$ , at which the system has the same free energy. This indicates a first-order

PT, with two phases coexisting at the PT temperature  $T^*$  that corresponds to the slope of this tangent. A schematic example of an  $S(E)$  curve featuring a convex intruder is shown in fig. 3.3. The convex behavior in  $S(E)$  translates to a loop-like feature in  $T(E)$ , which is also called a Gibbs-loop. The connection points of the tangent to  $S(E)$  correspond to the points of an equal-area-construction in  $T(E)$ , analogous to the well-known Maxwell-construction in the  $p$ - $V$  diagram. This method defines the coexistence region and the PT temperature and will be called *Gibbs-construction* in this work.

Additionally, one can use the microcanonical heat capacity  $c(E)$  to gain further insights into the phase behavior of the system.  $c(E)$  is related to the negative of the second derivative of the entropy with respect to energy, as shown in eq. (3.38). For a first-order PT, the loop region in  $T(E)$  results in an inflection point with a negative slope. This corresponds to a peak in  $c(E)$ , which has a negative value at the peak position. In contrast, a second-order PT is characterized by a monotonous change of the slope of the entropy. The corresponding region in  $T(E)$  does not feature a Gibbs-loop and the inflection point has a positive slope. Consequently, the maximum in  $c(E)$  for a second-order PT exhibits a positive value.

In this way, identification of maxima in the microcanonical heat capacity can be used to precisely determine the order of a PT. This approach offers greater accuracy compared to the evaluation of "sharpness" of peaks in  $C_V(T)$ .

### 3.4.3 Structural observables

In addition to the thermodynamic observables, structural observables were determined in the simulation runs. These observables give insight into the structure of the polymer chains and their interactions.

A universal and central structural observable is the tensor of gyration  $\mathbf{T}_g$ . It is a measure of the spatial extent of a polymer chain, and it is defined as

$$\mathbf{T}_g = \frac{1}{N_{\text{res}}} \sum_{i=1}^{N_{\text{res}}} (\vec{r}_i - \vec{r}_{\text{COM}}) \otimes (\vec{r}_i - \vec{r}_{\text{COM}}), \quad (3.44)$$

where  $N_{\text{res}}$  is again the number of residues in a single chain. Furthermore,  $\vec{r}_i$  is the position of the  $i$ -th bead in the chain,  $\vec{r}_{\text{COM}}$  is the position of the center of mass of the chain and  $\otimes$  is the tensor product. The tensor of gyration can be diagonalized to obtain the eigenvalues  $\lambda_1, \lambda_2, \lambda_3$ ,

which are sorted in descending order ( $\lambda_1 \geq \lambda_2 \geq \lambda_3$ ). From the eigenvalues, one can obtain the squared radius of gyration  $R_g^2$ :

$$R_g^2 = \frac{1}{N_{\text{res}}} \sum_{i=1}^3 \lambda_i^2. \quad (3.45)$$

Furthermore, one can obtain the relative shape anisotropy  $\kappa^2$ :

$$\kappa^2 = 1 - 3 \frac{\lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_3 \lambda_1}{(\lambda_1 + \lambda_2 + \lambda_3)^2}. \quad (3.46)$$

It is a measure for the symmetry and the dimensionality of the conformation of the polymer chain.  $\kappa^2$  is limited to the interval  $[0, 1]$ . It attains  $\kappa^2 = 0$  for a perfectly spherical conformation, and it becomes  $\kappa^2 = 1$  if all particles lie on a straight line.

A final descriptor one can obtain from  $\mathbf{T}_g$  is the asphericity parameter  $b$ . It measures the conformation's deviation from the spherical symmetry and is defined as:

$$b = \lambda_1 - \frac{1}{2}(\lambda_2 + \lambda_3). \quad (3.47)$$

Its interpretation becomes clear, when recalling that the eigenvalues  $\lambda_i$  are sorted in descending order.

Another important structural observable is the hydrodynamic radius  $R_h$ . It is a measure of the polymer chain's size in solution and is experimentally accessible from the diffusion coefficient  $D$ , the viscosity of the solvent  $\eta$  and the temperature  $T$ :  $R_h = k_B T / (6\pi\eta D)$ . For polymers in solution,  $R_h$  is also defined by

$$\frac{1}{R_h} = \frac{1}{N^2} \left\langle \sum_{i < j}^N \frac{1}{r_{ij}} \right\rangle, \quad (3.48)$$

where  $r_{ij}$  is the distance between the  $i$ -th and  $j$ -th bead in the chain and the brackets denote an ensemble average.

Hydrogen bond contact probability maps (HB maps) were used to identify characteristic HB patterns in the polymer chains. How a HB contact is formed is described in section 2.3. HB maps ( $\mathbf{P}_{ij}^{(\text{HB})}$ ) are  $(N \times N)$  matrices, where  $N = N_{\text{ch}} N_{\text{res}}$  is the total number of residues.  $N_{\text{ch}} = 2$  is the number of chains and  $N_{\text{res}}$  the number of residues per chain. A HB map shows the probability of a HB to be formed between the  $i^{\text{th}}$  NH and the  $j^{\text{th}}$  CO bead in the polymer chain:

$$\mathbf{P}_{ij}^{(\text{HB})}(U) = \frac{N_{ij}^{(\text{HB})}(U)}{H(U)}. \quad (3.49)$$

Here,  $N_{ij}^{(\text{HB})}(U)$  is the number times a HB was formed between residues the  $i^{\text{th}}$  NH and the  $j^{\text{th}}$  CO bead, if the configuration had a configurational energy  $U$  and  $H(U)$  is energy visitation histogram. Note, that the HB maps are not necessarily symmetric, because HBs form between different bead types (N and CO). Furthermore, one can calculate contact probability maps for side-chain interactions as well. They are calculated in the same way as the HB maps, but for side-chain beads instead of backbone beads.

The secondary structure of polymers can be identified using Ramachandran plots [73–76]. They display the distribution of the  $\Psi$  and  $\Phi$  angles of the residues in the polymer chain in a 2D map. For a more detailed introduction to Ramachandran plots, refer to section 2.1. Notably, certain secondary structures, such as  $\alpha$ -helices and  $\beta$ -sheets, restrict the  $\Psi$  and  $\Phi$  angles of the residues involved in these structures to specific regions within the Ramachandran plot (see fig. 2.2). Amino acids with torsion angles in the range of  $\Psi \in [-100^\circ, 45^\circ]$  and  $\Phi \in [-180^\circ, 0^\circ]$  are considered to be in the  $\alpha$ -helical region. Amino acids with torsion angles in the range of  $\Psi \in [45^\circ, 225^\circ]$  and  $\Phi \in [-180^\circ, -45^\circ]$  are considered to be in the  $\beta$ -sheet region. This allows for the identification of the secondary structure to which a residue belongs. Böker et

al. used Ramachandran plots of homopolymer systems to identify the correct model version of PRIME20 [53, 61] by comparing the obtained Ramachandran plots with the proteins known secondary structure. In this work heteropolymer systems were investigated, which have diverse secondary structure motifs distributed along the polymer chain. Given the complexity of these systems, examining the Ramachandran plots of the entire chain proves impractical. Instead, a more effective approach was chosen that takes the Ramachandran plots of the individual residues, identifies their corresponding secondary structure and then calculates the probability distribution of these structures along the chain.

## Chapter 4

# Dimerization of Polyglutamine

The model and simulation method described in chapters 2 and 3 were used to investigate three different peptide systems: polyglutamine (PolyQ), Amyloid beta ( $A\beta$ ) and parathyroid hormone (PTH). This chapter focuses on the results obtained from the PolyQ systems, while the findings for the other two peptides are discussed in chapters 5 and 6. The primary results of the PolyQ investigation have been previously published [56]. In this chapter, these main findings are summarized, expanded upon and discussed within a broader context.

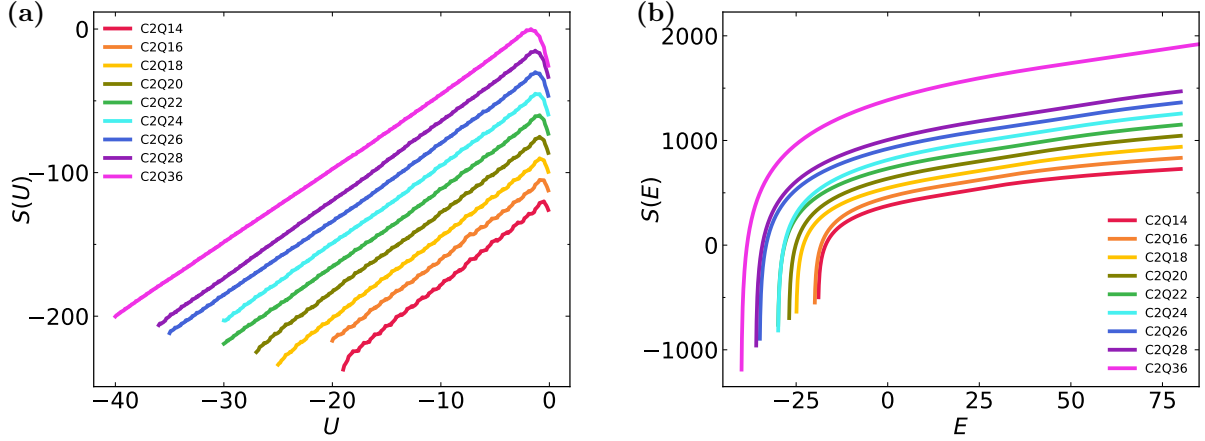
Additionally, this chapter will serve as an introduction to the analysis methods used throughout this thesis. The methods include a thermodynamic analysis, which starts with the configurational density of states (DOS) obtained from the SAMC simulations and leads to an identification of the systems phase transitions. Furthermore, an analysis of the configurations in the different phases will be carried out. This includes using structural observables such as the tensor of gyration and hydrogen bond contact probability matrices. The chapter will conclude with a discussion of the results and their implications for the aggregation of PolyQ peptides. The analysis methods used in this chapter will largely be applied to the  $A\beta$  and PTH systems in the subsequent chapters as well.

Simulations were performed on PolyQ systems of two chains of identical length. The chain length ( $N_{\text{res}}$ ) here refers to the degree of polymerization i.e., the number of amino acid residues in the chain. Nine different chain lengths were studied:  $N_{\text{res}} \in (14, 16, 18, 20, 22, 24, 26, 28, 36)$ . The side length of the simulation box was set to  $L = 112.5$  or  $L = 150$  depending on the chain length (see section 3.3.2 and in particular table 3.2). These settings correspond to millimolar concentrations, which is close to in vitro experimental conditions for PolyQ aggregation. Throughout this chapter, the different PolyQ systems will be designated by the code 'CXQY', where X denotes the number of chains and Y represents the number of residues per chain. For example, C2Q24 refers to a system of 2 chains, each with 24 Gln residues.

### 4.1 Thermodynamic analysis

As described in chapter 3, the SAMC algorithm produces an approximation of the microcanonical density of states  $g(U)$ . More precisely, it produces an approximation of  $\ln g(U)$  which is equal to the entropy  $S(U)$  when setting the Boltzmann constant  $k_B = 1$ . It is important to note, that the approximated  $S(U)$  is known only up to an additive constant. However, since the analysis of phase transitions requires the calculation the derivatives of  $S(U)$ , the additive constant has no impact on the here performed analysis. It can thus be chosen freely. The entropies of the nine different PolyQ systems are shown in fig. 4.1(a). The shift along the y-axis between the different systems is chosen to better differentiate them. This shift is permissible here due to the aforementioned unknown additive constant.

Looking at the behavior over the entire energy range, the entropies all exhibit roughly the same ascending slope. This behavior is to be expected, since the entropy is a measure for the

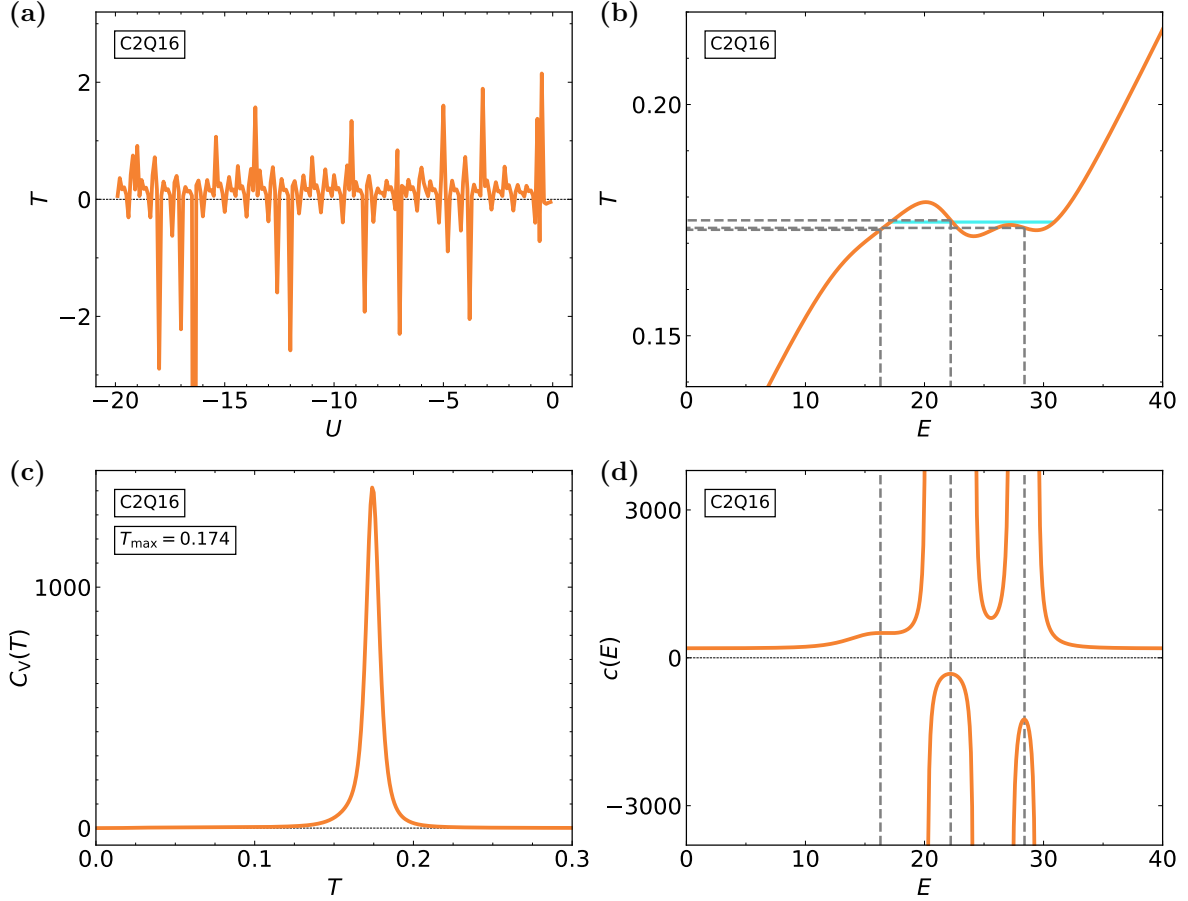


**Figure 4.1:** Configuration and total-energy microcanonical entropy of the 9 polyglutamine systems of varying chain length. The entropies are shown in (a) as a function of the potential energy  $U$  and in (b) as a function of the total energy  $E$ . The different systems are shifted along the y-axis for better visibility as they are only determined by the SAMC algorithm up to an additive constant.

number of microstates corresponding to a given energy  $U$ , and the low energy regime imposes higher restrictions on the systems conformational freedom. The lower bound of the displayed  $S(U)$  is defined by the lowest energy at which the simulations of the corresponding system converged.

Another notable observation is that the maximum of the entropy occurs at an energy below zero, rather than at the system's maximum potential energy. This behavior of  $S(U)$  can be attributed to the geometric properties of the polymer chain in the PRIME20 model. For a configuration to have an energy of zero, it has to be in a state where neither HB nor SC interactions are present. For the displayed PolyQ systems, such configurations are possible. However, only small alterations to these configurations can cause SC beads to experience each others square-well attraction (eq. (2.3)). This makes these *no-contact* configurations less probable compared to those with at least some degree of SC contact. Consequently, the maximum in entropy, corresponding to the highest number of microstates, occurs at an energy below zero. Its exact value depends on the chain length, with longer chains having a maximum at lower energies compared to shorter chains.

Another feature of the entropies in fig. 4.1(a) that springs to attention is their oscillating behavior. This is most prominent in the smaller systems, such as C2Q14 and C2Q16. The oscillations are the result of the discretization of energy space and the two different energy scales in the PRIME20 model. Consider the two energy intervals  $E_{[-]} = (-1 - \delta E; -1]$  and  $E_{[+]} = (-1; -1 + \delta E]$ . Here  $\delta E = 0.1$  is the width of the energy bin used in the simulation. The energy gain of a HB in the PRIME20 model is defined as  $E_{\text{HB}} = -1$  (see section 2.3.1). Thus, a configuration which has one established HB falls into the energy bin  $E_{[-]}$ , just below  $E_{[+]}$ . Configurations can only belong to  $E_{[+]}$  if they do not have any established HBs. Instead, SC contacts have to be formed in order for the configuration to have an energy within  $E_{[+]}$ . The energy gain of SC interactions for PolyQ chains in the PRIME20 model is  $E_{\text{SC}} = -0.08$ . Consequently, 12 SC contacts have to be formed in order to reach the energy  $E = -0.96 \in E_{[+]}$ . It is obvious that the formation of 12 SC contacts imposes greater steric restrictions on the possible configurations compared to the formation of a single HB. This results in a lower entropy in  $E_{[+]}$  compared to  $E_{[-]}$ . This holds especially true for systems of shorter chains, where the number of available SC beads is smaller when compared to systems of longer chains. The oscillations in the entropy are thus a direct consequence of the two different energy scales (for HBs and SCs) in the PRIME20 model. The oscillations in  $S(U)$  have a significant impact on its derivatives. The temperature curves  $T(U)$ , which are the inverses of the first derivatives of



**Figure 4.2:** Plots relevant to the thermodynamic analysis of the C2Q16 system. (a) Temperature  $T(U)$  derived from the configurational entropy  $S(U)$ , (b) temperature  $T(E)$  derived from the full-energy entropy  $S(E)$ , (c) canonical heat capacity  $c_V(T)$  and (d) microcanonical heat capacity  $c(E)$ . In (b) and (d) the gray dashed lines mark the energies of the phase transitions.

$S(U)$ , exhibit large peaks in both positive and negative direction. For example, the temperature derived from the configurational entropy of the C2Q16 system is shown in fig. 4.2(a). Negative values for the microcanonical temperature are not uncommon for finite systems and simulations of the same. Their physical significance has been widely discussed in the literature [121, 142, 156, 157]. However, the extreme fluctuations in the temperatures of the PolyQ systems are unusual and not easily interpretable.

Therefore, the  $S(U)$  graphs are smoothened by including the contribution of the kinetic part of the energy. This is achieved using a convolution, according to eq. (3.33). This conversion is further motivated by the fact, that the microcanonical ensemble is defined in phase space, not only configuration space, as has already been discussed in section 3.4.

The full-energy microcanonical entropies  $S(E)$  for all PolyQ system sizes are shown in fig. 4.1(b). The  $S(E)$  curves are monotonous and smooth, unlike the  $S(U)$  curves. Again, their minimum is defined by the lowest energy at which the simulations of the corresponding system converged at. The upper energy bound is infinity. When going to large energies, the DOS will tend towards the DOS of the ideal gas as the kinetic contributions will dominate. In order to interpret the shape of  $S(E)$ , its derivatives are calculated. The temperature  $T(E)$  and the heat capacity  $c(E)$  are directly related to the first and second derivatives of  $S(E)$  (see eqs. (3.37) and (3.38)). Again, the C2Q16 system is used as an example.  $T(E)$  is shown in fig. 4.2(b) and  $c(E)$  in fig. 4.2(d).

Inflection points in  $T(E)$  indicate pseudo phase transitions (PTs) in the system [158–160].

The prefix "pseudo" arises, because in statistical mechanics, PTs are only defined in the thermodynamic limit. However, the here presented results are for simulations of finite systems. Nevertheless, in the remainder of this work the prefix will be dropped for better readability. Following the Ehrenfest classification [161], a PT is classified as first-order if the slope of the temperature  $T(E)$  at the corresponding inflection point at  $E = E^*$  is negative. Therefore, the temperature curve is non-monotonic. Since  $c(E)$  is the inverse of the first derivative of  $T(E)$ , an inflection point in  $T(E)$  corresponds to a local maximum in  $c(E)$ . If the slope of  $T(E)$  at  $E = E^*$  is negative, then  $c(E^*) < 0$  (*negative* maximum). In this scenario, system is in a "mixed-phase regime" in which some parts of the system have completed the transition and others have not. Conversely, if the slope of  $T(E)$  at  $E = E^*$  is positive, the PT is of second order and  $c(E^*) > 0$  (*positive* maximum). A second-order PT is also called a *continuous* PT.

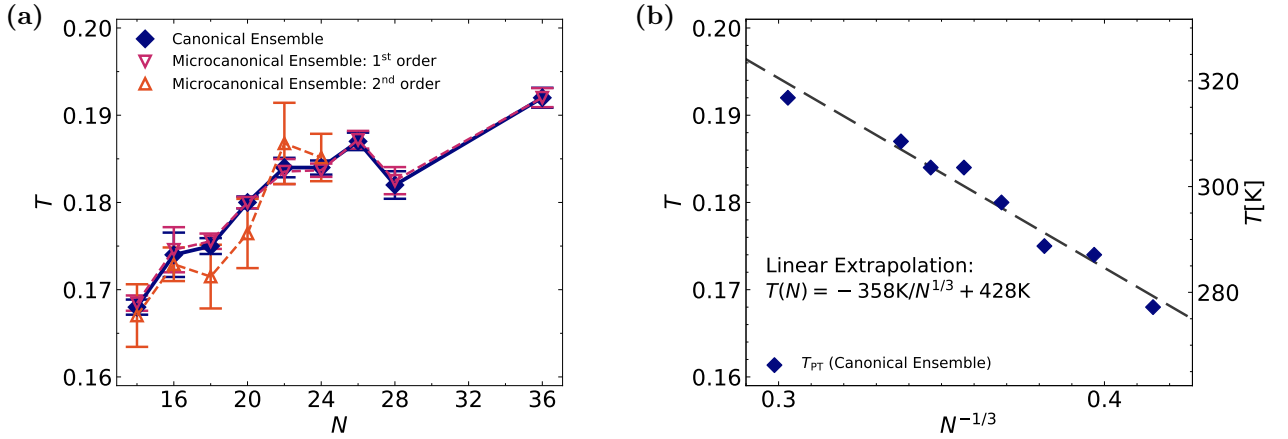
In fig. 4.2(d), three local maxima are visible in the heat capacity, two with negative and one with positive value. The energies of these maxima are marked with gray dashed lines and their positions are carried over to the temperature curve in fig. 4.2(b). All three maxima correspond to roughly the same temperature. The two negative maxima in  $c(E)$  align with inflection points in  $T(E)$  that are part of the same large-scale oscillation. They occur at the same temperature and are indicative of the same transition. The temperature of this unifying first-order PT can be determined using a Gibbs-construction (see section 3.4.2). This involves placing a horizontal line in the loop region of the  $T(E)$  graph such that the areas enclosed by the line and the graph above and below the line are equal. In fig. 4.2(b) this results in the blue line. The position of the horizontal line on the  $T$ -axis gives the transition temperature. Although the two negative maxima in  $c(E)$  merge into a single first-order transition temperature through the Maxwell construction, the two inflection points in  $T(E)$  indicate that there are indeed two first-order PTs occurring at this temperature. As will be shown later, these transitions correspond the folding and the aggregation transition.

The  $T(E)$  graphs for the other PolyQ systems all show an oscillating behavior similar to the C2Q16 system. They can be found in appendix B in fig. B.4. The first-order PT temperatures for all systems are determined using the Gibbs construction. All systems show two first-order peaks in their heat capacities. The only exceptions are C2Q20, C2Q28 and C2Q36. In the C2Q20 system a shoulder of the "missing" maxima is present. However, its peak gets superimposed with the neighboring peak. The C2Q28 system shows three negative maxima in  $c(E)$ , which again are part of the same large-scale oscillation in  $T(E)$ . The three maxima in  $c(E)$  translate to approximately the same temperature. In the C2Q36 system, no distinction can be made between two separate first-order transition signatures in the  $c(E)$  curve, as just one negative maximum is visible. The transition temperatures found in the microcanonical ensemble for all systems are listed in table 4.1. Next to the first order PT temperatures, the table also shows the temperatures of the second order PTs found in the microcanonical ensemble. Short chain systems from chain length 14 to 24 show signatures of second order PT in their  $c(E)$  curves. Systems of longer chains do not. These second order PTs occur at lower energies than the first-order PTs. However, they again translate to roughly the same transition temperature. Additionally, the C2Q14 systems exhibits a second order PT at a low temperature around  $T \approx 0.06$ . This indicates a restructuring of the configurations in the system after the folding and aggregation, which will be explored in further detail later in this chapter. The other systems do not show this low-energy second order PT peak in  $c(E)$ .

In order to support the findings in the microcanonical ensemble the heat capacities in the canonical ensemble are calculated. The canonical partition function  $Z(N, V, T)$  is obtained from  $g(E)$  as described by eq. (3.20). Using  $Z(N, V, T)$ , the canonical heat capacity  $C_V(T)$  can be calculated using eq. (3.40). Maxima in  $C_V(T)$  again correspond to PTs. In infinite systems the heat capacity would diverge at the temperature of a first order PT. However, in finite systems this instead translates to sharp peaks in  $C_V(T)$  at this temperature. In relatively small systems like the one at hand, a second order PT produces a broad maximum in  $C_V(T)$ . The heat capacity

**Table 4.1:** Phase transition temperatures  $T^*$  for all investigated PolyQ systems as obtained from microcanonical and canonical analysis of the entropy  $S(E)$ . First order phase transition temperatures in the microcanonical ensemble  $T_{\text{micro}}^*$  are determined using the Gibbs construction and second order phase transition temperatures by finding local maxima with positive value in the microcanonical heat capacity.  $T_{\text{can}}^*$  is derived from the maxima in the canonical heat capacity.

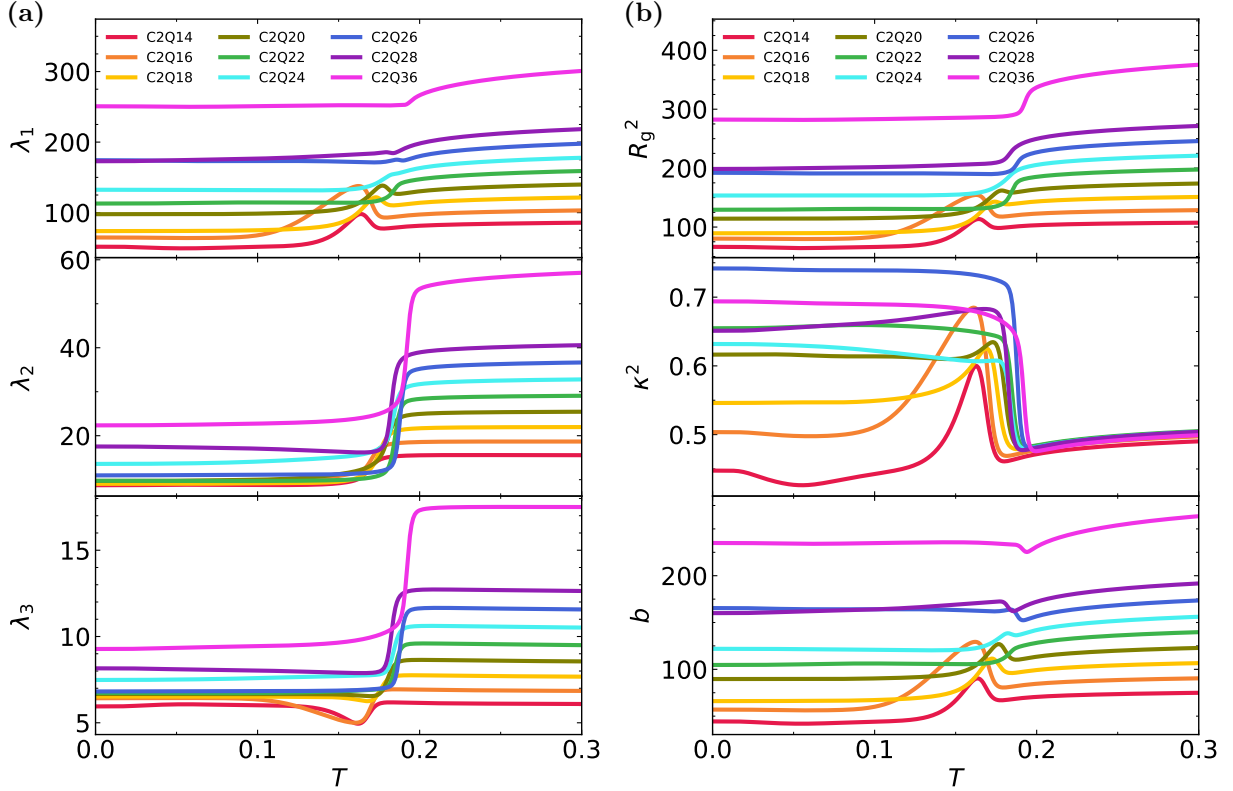
$N$	$T_{\text{micro}}^*$		$T_{\text{can}}^*$
	1 <sup>st</sup> order	2 <sup>nd</sup> order	
14		$0.0601 \pm 0.0066$	$0.0600 \pm 0.0009$
16	$0.16846 \pm 0.00088$	$0.1670 \pm 0.0036$	$0.1680 \pm 0.0009$
18	$0.17555 \pm 0.00088$	$0.1715 \pm 0.0037$	$0.1750 \pm 0.0009$
20	$0.17998 \pm 0.00065$	$0.1764 \pm 0.0040$	$0.1800 \pm 0.0007$
22	$0.1835 \pm 0.0014$	$0.1868 \pm 0.0047$	$0.1840 \pm 0.0011$
24	$0.18369 \pm 0.00076$	$0.1852 \pm 0.0027$	$0.1840 \pm 0.0008$
26	$0.1872 \pm 0.0010$		$0.1870 \pm 0.0010$
28	$0.1825 \pm 0.0016$		$0.1820 \pm 0.0016$
36	$0.1920 \pm 0.0011$		$0.1920 \pm 0.0012$



**Figure 4.3:** Transition temperatures of the polyglutamine systems, determined from canonical and microcanonical analysis, as a function of chain length. On the left (a) results comparing the different methods are shown. On the right (b) the canonical transition temperatures are plotted against  $N^{-1/3}$  to show the scaling behavior. The dashed line is a linear extrapolation to infinite chain length. The origin of the temperature scale in Kelvin on the right y-axis of (b) is discussed in section 4.3. The figures are adapted from Lauer et al. [56].

for the C2Q16 system is shown in fig. 4.2(c). It exhibits a distinct peak at a temperature around  $T \approx 0.175$ . The heat capacities of the other chain lengths also exhibit such first order peaks. The PT temperatures found in the canonical ensemble for all systems are listed in table 4.1. The table shows that the first order PT temperatures found in the canonical ensemble are in good agreement with the first order PT temperatures found in the microcanonical ensemble. This is illustrated in fig. 4.3(a), where the PT temperatures identified in the different ensembles are plotted vs. the chain length. The low-temperature second-order PT in the C2Q14 system is not shown in fig. 4.3(a) but as one can see in table 4.1 it is also found in the canonical ensemble.

The deviations between the different ensembles arise from finite size effects in the simulation environment and should disappear in the thermodynamic limit. The PT temperatures increase with the chain length. Only the C2Q28 system deviates from this behavior as it has a lower PT temperature than the C2Q26 system. This is likely due to the convergence issues of the



**Figure 4.4:** Quantities derived from the tensor of gyration as a function of temperature. (a) Eigenvalues  $\lambda_i$  of the tensor of gyration  $\mathbf{T}_g$  for all PolyQ systems. (b) Radius of gyration  $R_g$ , relative shape anisotropy  $\kappa^2$  and asphericity  $\Delta$  for all PolyQ systems. Units:  $[\lambda_i] = \text{\AA}$ ,  $[R_g] = \text{\AA}^2$ ,  $[\kappa^2] = 1$ ,  $[\Delta] = 1$ .

simulations of the C2Q28 system. Because the simulation did not converge for low enough energies for this system size, the derived PT temperatures are too low. In the thermodynamic limit, aggregation transitions are of first-order and thus finite size corrections scale as  $N^{-1/3}$  [162, 163]. Folding transitions typically are of first-order as well, which suggests using the same scaling behavior. This is applied in fig. 4.3(b). The data is fitted with a power law function of the form  $T^*(N) = T_\infty^* + aN^{-1/3}$ . The fit is in good agreement with the data. In the scaling analysis, the C2Q28 system is excluded due to the aforementioned convergence issues.

Using the analysis of  $\ln g(U)$  and the subsequent calculation of the temperature and heat capacity, the PT temperatures of the PolyQ systems were identified. The next step is to analyze the configurations of the systems in the different phases, thereby confirming the previous attribution of the PTs to the folding and aggregation processes.

## 4.2 Structural analysis

### 4.2.1 Tensor of gyration

The tensor of gyration  $\mathbf{T}_g$  is a measure of the spatial distribution of the particles in a system. It is defined in the previous chapter in eq. (3.44). When diagonalizing the tensor, the eigenvalues  $\lambda_i$  with  $i \in (1, 2, 3)$  can be obtained. They are sorted in ascending order:  $\lambda_1 > \lambda_2 > \lambda_3$ . The eigenvalues describe the extension of the system along the principal axes. Their averaged values obtained from the simulations are shown in fig. 4.4(a) for all systems. When cooling,  $\lambda_2$  and  $\lambda_3$  show a significant drop at a temperature that coincides with the PT temperatures  $T^*$  obtained from the heat capacity analysis. This drop is present in all systems, although most pronounced in systems of longer chains. It separates a high-temperature phase from a low-temperature

phase. In contrast, the drop in  $\lambda_1$  is smaller when compared to  $\lambda_2$  and  $\lambda_3$ , yet the drop is still present. Furthermore, small systems with  $N_{\text{res}} \in (14, 16, 18, 20)$  show a slight increase in  $\lambda_1$  at  $T^*$ , which indicates an elongation of the chains. When cooling further,  $\lambda_1$  decreases again to values smaller than above the transition. The behavior of  $\lambda_1$  is mirrored in  $\lambda_3$ , where in small systems the eigenvalue drops at  $T^*$  before increasing again until it stabilizes at chain length specific value at low temperatures.

In small systems with  $N_{\text{res}} \in (14, 16, 18, 20, 22)$ , the two small eigenvalues group around similar values at low temperatures:  $\lambda_2 \approx 9$  and  $\lambda_3 \approx 7$ . Additionally, the C2Q26 system also reaches these values for  $\lambda_2$  and  $\lambda_3$ . In  $\lambda_3$  no such grouping can be observed.

In the C2Q14 system a low temperature transition at  $T_{14,\text{low}}^* \approx 0.06$  has been identified in the thermodynamic analysis. At  $T_{14,\text{low}}^*$  no significant change is observed in the eigenvalues of  $\mathbf{T}_g$ . However, it is worth noting that  $T_{14,\text{low}}^*$  lies in the regime, where  $\lambda_1$  and  $\lambda_3$  maintain their stable low temperature values.

The behavior of the  $\mathbf{T}_g$  eigenvalues of the short chain systems suggests an elongation of the chains at  $T^*$ , followed by a collapse at lower temperatures. This observation could indicate an initial aggregation of stretched chains into a  $\beta$ -sheet conformation, which then collapses into aggregated two  $\beta$ -hairpins. For systems of longer chains, no elongation at  $T^*$  can be observed. Instead, the signatures in the  $\mathbf{T}_g$  eigenvalues are consistent with the formation of  $\beta$ -hairpins at  $T^*$ . Such a chain length dependent behavior has been suggested previously by Chen et al. [164]. They found that for short chains ( $N_{\text{res}} \leq 20$ ) extended structures are favored, while longer chains ( $N_{\text{res}} = 30$ ) prefer to form  $\beta$ -hairpin structures.

This difference between long and short chain systems can also be observed in the radius of gyration  $R_g$ . The  $R_g$  describes the average distance of the particles from the center of mass of the system and can be calculated by summing over the squares of the eigenvalues of  $\mathbf{T}_g$  (see eq. (3.45)).  $R_g$  is displayed for all PolyQ systems in fig. 4.4(b) in the top graph. The  $R_g$  of the short chain systems shows a significant increase at  $T^*$ , which is indicative of an elongation of the chains. The  $R_g$  of the long chain systems do not show this behavior. Instead, the  $R_g$  of the long chain systems decrease at  $T^*$ , which suggests the formation of more compact configurations at the PT.

Another quantity derived of  $\mathbf{T}_g$  is the relative shape anisotropy  $\kappa^2$  (see eq. (3.46)). It measures symmetry and dimensionality of a conformation. For an ideal rod-like object  $\kappa^2 = 1$ , for a perfect sphere  $\kappa^2 = 0$ . The shape anisotropy is shown in the middle of fig. 4.4(b). For  $T > T^*$ ,  $\kappa^2$  has the same value for all chain lengths around  $\kappa^2 \approx 0.5$ . This indicates that in this high-temperature phase, the configurations in all systems are neither rod like, nor spherical. They all have a similar shape and are in the random coil state. At  $T \approx T^*$ ,  $\kappa^2$  increases for all systems. This corresponds to the formation of more rod-like configurations, like  $\beta$ -sheets. In systems of  $N_{\text{res}} \in (14, 16, 18, 20)$ ,  $\kappa^2$  decreases again when cooling further, until it reaches a system size specific value at low temperatures. In the long chain systems on the other hand,  $\kappa^2$  remains at a high value at low  $T$  after the increase at  $T^*$ . The behavior of  $\kappa^2$  is consistent with the different behavior of the  $\mathbf{T}_g$  eigenvalues in the short and long chain systems. The chains in the short chain system elongate at  $T^*$ , followed by a collapse at low temperatures. The long chain systems show an elongation indicative of  $\beta$ -hairpin formation at  $T^*$ , which persists at low temperatures. Additionally,  $\kappa^2$  allows some insight into the low temperature transition of the C2Q14 system. As C2Q14 shares the same behavior around  $T^*$  with the other short chain systems,  $\kappa^2$  decreases again after the initial increase, when going from large to small  $T$ . In the low temperature regime,  $T_{14,\text{low}}^*$  sits in the global minimum of  $\kappa^2$ , where the conformations are most symmetric and spherical. Below  $T_{14,\text{low}}^*$ ,  $\kappa^2$  increases again as the system restructures into a more rod-like shape.

The asphericity parameter  $b$ , the third quantity derived from  $\mathbf{T}_g$ , is shown in the bottom graph of fig. 4.4(b). This parameter measures the configuration's deviation from spherical symmetry. As a linear combination of the  $\mathbf{T}_g$  eigenvalues (see eq. (3.47)), its behavior is similar to

that of  $\lambda_1$ . In short chain systems, an increase in  $b$  is observed at  $T^*$ , followed by a decrease at lower temperatures. This again is consistent with the picture of elongation and collapse. In contrast, long chain systems shows only a slight decrease in the asphericity parameter at  $T^*$ . In systems with  $N_{\text{res}} \in (26, 28, 36)$ , the decrease at  $T^*$  is followed by a small increase. Even though the information provided by the asphericity parameter is somewhat redundant, it validates the findings of the other quantities derived from  $\mathbf{T}_g$  discussed above.

### 4.2.2 Hydrogen bond contact probability maps

Hydrogen bond contact probability maps (HB maps) are investigated in order to identify characteristic configurations that are formed below the transition temperature. HB maps are  $N \times N$  matrices, where  $N = N_{\text{ch}}N_{\text{res}}$  is the total number of residues, with  $N_{\text{ch}} = 2$  being the number of chains and  $N_{\text{res}}$  being the number of residues per chain. Each cell  $(i, j)$  represents a possible HB between bead  $\text{NH}_i$  and  $\text{CO}_j$ . The darker the color in the cell, the more likely a HB is formed between the corresponding beads. For the probability calculation see eq. (3.49). The residues in the 2 chains are labeled in a consecutive manner, such that the NH and CO beads in chain 1 are labeled from 1 to  $N_{\text{res}}$  and the NH and CO beads in chain 2 are labeled from  $(N_{\text{res}} + 1)$  to  $2N_{\text{res}}$ . This allows for the simultaneous display of intra and inter-molecular contact. Looking at the HB maps in fig. 4.5, the top left and bottom right quadrants show the inter-molecular HB contacts. They will be called *inter-quadrants* from here on out. The top right and bottom left quadrants show the intra-molecular HB contacts and will be called *intra-quadrants*. A feature common to all intra-quadrants is a zero-probability area stretching from bottom left to top right. This results in an always-present white diagonal in the intra-quadrants. This is due to the rules of HB formation in the PRIME20 model, which does not allow HB to form between beads from residues of the same peptide chain that are separated by less than 4 other residues.

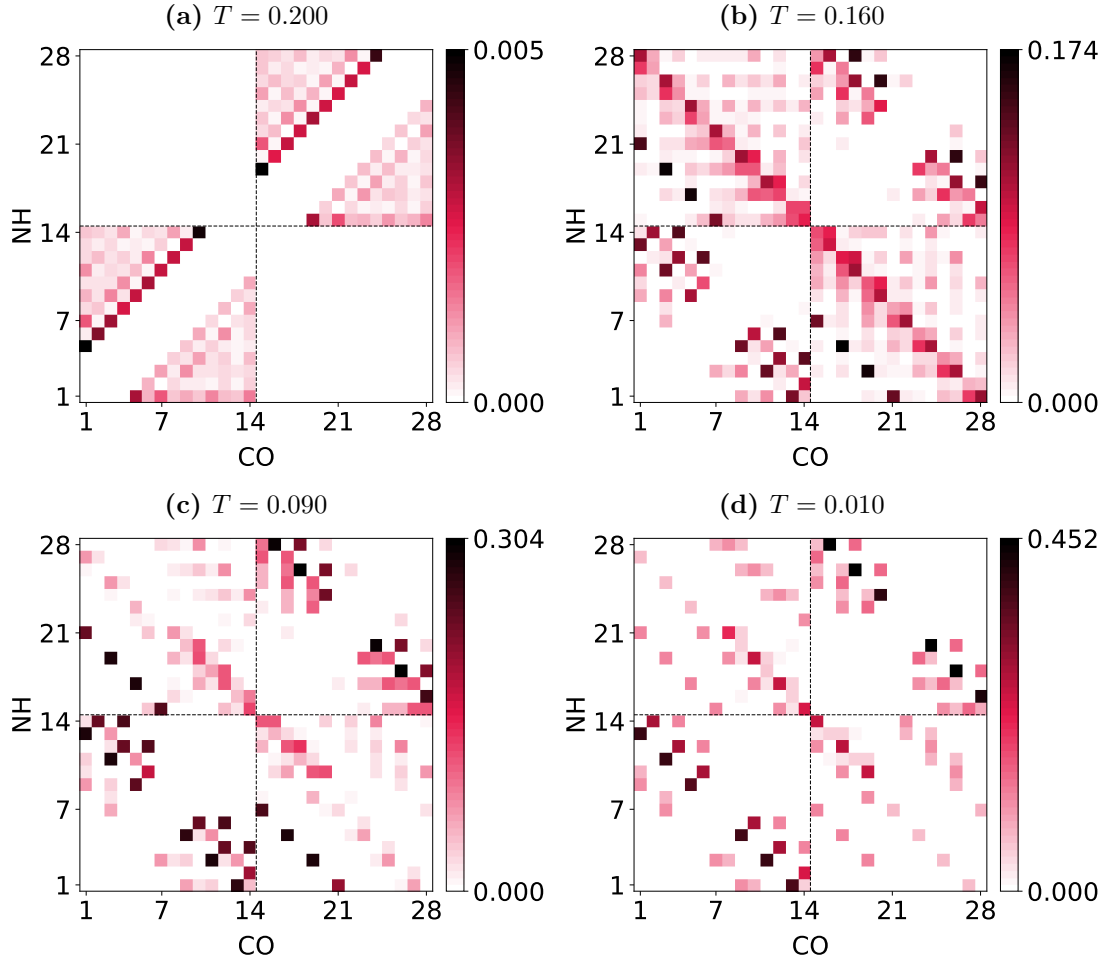
#### C2Q14

Figure 4.5 displays the HB maps of the C2Q14 system at four different temperatures. The first map shows the system at a temperature well above the high-temperature transition,  $T_{14,\text{high}}^* = 0.168$  (fig. 4.5(a)). The second map corresponds to a temperature just below  $T_{14,\text{high}}^*$  (fig. 4.5(b)). The third map depicts the system at an intermediate temperature, between the high- and low-temperature transitions,  $T_{14,\text{high}}^*$  and  $T_{14,\text{low}}^*$  (fig. 4.5(c)). Finally, the fourth map illustrates the system at a temperature below the low-temperature transition,  $T_{14,\text{low}}^* = 0.060$  (fig. 4.5(d)).

Looking at the HB map at  $T = 0.200$  in fig. 4.5(a), the probability of any HB forming is very low. The polymer chains are in a random coil state. The only HB signatures that are visible are in the intra-quadrants. The beads with the highest probability of a HB contact are close to the turn region with  $|i - j| = 4$ , where  $i$  and  $j$  are the indices of the bead's residues. This seems reasonable, considering that only minor bending of the chain is required for the beads of residues  $i$  and  $j$  to be in appropriate proximity to each other.

At  $T = 0.160$  (fig. 4.5(b)), just below  $T_{14,\text{high}}^*$ , the probability of HB formation is increased, compared to the high-temperature state in fig. 4.5(a). Intra as well as inter-molecular HBs are formed which confirms the previous assignment of the  $C_V(T)$  peak at  $T_{14,\text{high}}^*$  to a folding and aggregation transition. The most notable feature in the HB map in this collapsed state are the diagonal signatures with descending slopes. They are indicative of anti-parallel alignment of the chains and  $\beta$ -sheet formation. As described in section 2.1,  $\beta$ -sheets are stabilized by HBs between the NH and CO beads of the backbone. Subsequent residues that take part in the HB pattern of the  $\beta$ -sheet are next-nearest neighbors. If HBs are formed between the residue pair  $(i, j)$ , then the next residues in the  $\beta$ -sheet to form HBs are  $(i + 2, j - 2)$ ,  $(i + 4, j - 4)$ , and so on. This results in the diagonal signatures in the HB map. Since every residue has a NH and a CO bead, 2 HBs can form between residues  $i$  and  $j$ :  $(\text{NH}_i, \text{CO}_j)$  and  $(\text{CO}_i, \text{NH}_j)$ .

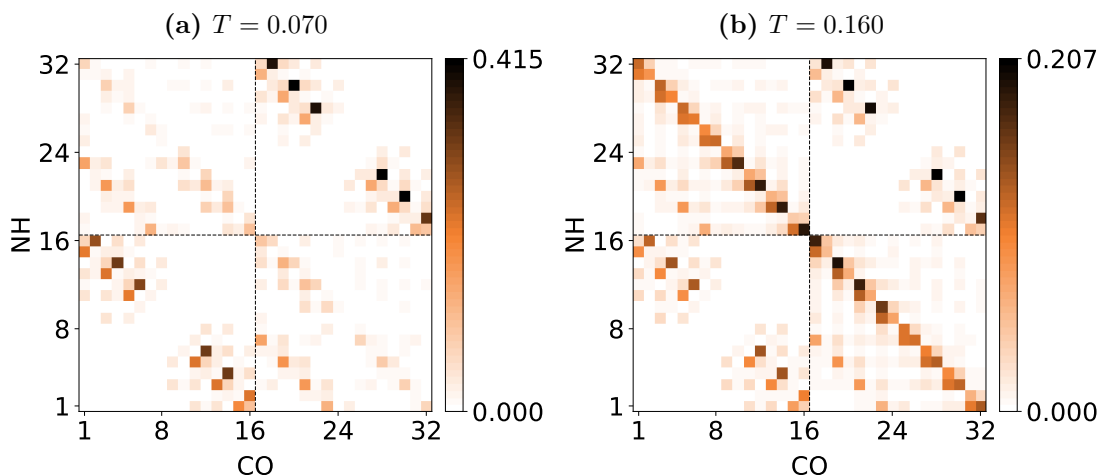
In the intra-quadrants, the diagonal signatures indicate the formation of  $\beta$ -hairpins. Two



**Figure 4.5:** Hydrogen bond contact matrices for chain length  $N = 14$  at four different temperatures. Each cell  $(i, j)$  corresponds to a possible contact between the NH bead of residue  $i$  and the CO bead of residue  $j$ . The color indicates the probability of a hydrogen bond forming between the beads, according to the color scale on the right of the maps. The colors are scaled to the highest value in the map. Therefore, equal colors in different maps do not correspond to the same contact probability.

$\beta$ -sheet segments that form HBs between each other are connected by a turn consisting of a varying number of residues and no HBs. The minimal number of residues in the turn is again defined by the PRIME20 model's HB criteria, which dictates, that the closest possible HB is  $(i, i + 4)$ . The most probable  $\beta$ -hairpins occupy the main-diagonals in the quadrants. Here, the turn is in the middle of the chain and the configurations therefore have the highest number of intra-HBs. These  $\beta$ -hairpins then aggregate into proto-fibrils. They consist of 2  $\beta$ -hairpins with established HBs between each other, creating an anti-parallel inter-molecular  $\beta$ -sheet. Examples of such structures can be seen in figs. 4.8(a) and 4.8(b). In the inter-quadrants, the aggregated  $\beta$ -hairpin structures are visible as diagonals of 3 or 4 HBs that span half the diagonal width of the quadrant. In  $\beta$ -sheets, the orientation of the backbone alternates between neighboring residues. Thus, in the chain segments that belong to both  $\beta$ -hairpin and aggregated  $\beta$ -sheet, if residue  $i$  is part of the  $\beta$ -hairpin, then the neighboring residue  $(i + 1)$  is part of the aggregated  $\beta$ -sheet. This fits with the HB maps of the aggregated state (figs. 4.5(b) to 4.5(d)): for every  $\beta$ -hairpin signature in an intra-quadrant there is a matching  $\beta$ -sheet signature in an inter-quadrant.

Next to the signatures of aggregated  $\beta$ -hairpins, signatures of another aggregated structure are visible in fig. 4.5(b). In the inter-quadrants, a diagonal signature of high contact probability that spans over the entire main-diagonal is visible. This signature belongs to configurations in which both chains form extended  $\beta$ -strands and form a  $\beta$ -sheet consisting of both entire PolyQ



**Figure 4.6:** HB matrices of C2Q16 at temperatures below the transition temperature, corresponding to the two ordered states.

chains in anti-parallel alignment. The PolyQ chains form 7 pairs of inter-molecular HBs to create a whole-chain  $\beta$ -sheet structure. Figure 4.8(c) shows such a whole-chain  $\beta$ -sheet configuration of the C2Q14 system.

When going to  $T = 0.090$  (fig. 4.5(c)), the probability of the formed HBs increases. Furthermore, the whole-chain  $\beta$ -sheet signatures that were present at  $T = 0.160$  disappear. The half-diagonal signatures on the other hand remain. This indicates that the aggregated  $\beta$ -hairpins are still present, whereas the whole-chain  $\beta$ -sheet structures are not stable anymore at this temperature.

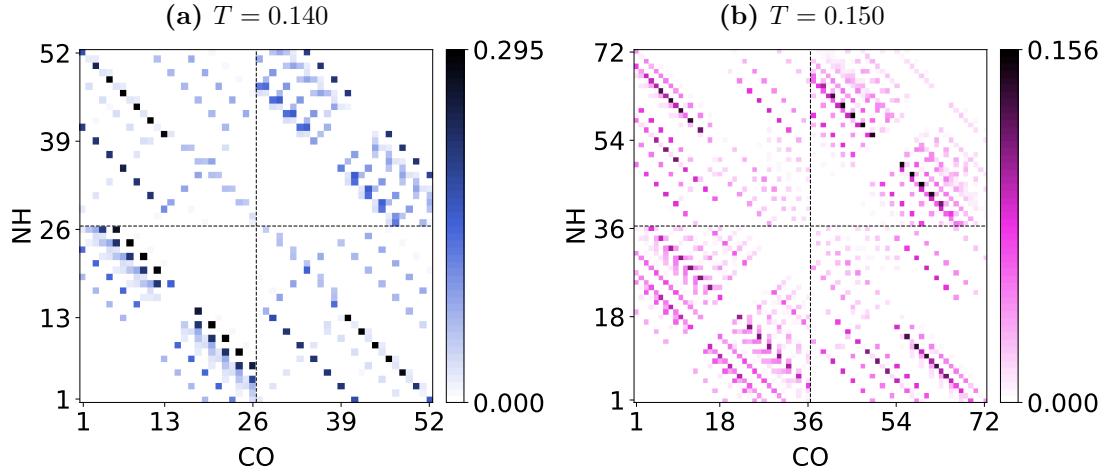
At  $T = 0.010$  (fig. 4.5(d)), the probability of HB formation is further increased. Additionally, HBs that belong to off-center diagonals in the intra-quadrants become less probable, compared to those in the main-diagonal. The  $\beta$ -hairpins corresponding to the off-center diagonals are less optimal than the centered  $\beta$ -hairpins. This can be explained by the decreased number of possible intra-molecular HBs if the turn in the  $\beta$ -hairpin is not centered in the chain. The configurations that are most probable at this temperature are those that have 6 intra-molecular HBs in each chain and 7 inter-molecular HBs (fig. 4.8(b)).

## C2Q16

The C2Q16 system exhibits similar behavior to the C2Q14 system. Two HB maps of the C2Q16 system at different temperatures are shown in fig. 4.6. At  $T = 0.160$ , just below the transition temperature  $T_{16}^* = 0.174$ , the HB map displays signatures of two types of configurations. On the one hand, high-probability signatures spanning the entire main-diagonal of the inter-quadrants indicate the presence  $\beta$ -sheet structures consisting of whole-chains. On the other hand,  $\beta$ -hairpin signatures are visible as full-diagonals in the intra-quadrants. These have corresponding  $\beta$ -sheet half-diagonal patterns in the inter-quadrants, suggesting aggregated  $\beta$ -hairpins. An example of such a structure is shown in fig. 4.8(d). When cooling to  $T = 0.070$ , well below the transition temperature, the whole-chain  $\beta$ -sheet signatures disappear, while the  $\beta$ -hairpin signatures persist (see fig. 4.6(a)).

## C2Q26 and C2Q36

In system of longer chains, the behavior differs significantly. As examples, the HB maps for C2Q26 and C2Q36 are shown in fig. 4.7. Both HB maps are taken at temperatures below their systems transition temperature, as identified from the canonical heat capacity analysis. Unlike the C2Q14 and C2Q16 systems, the HB maps of C2Q26 and C2Q36 do not exhibit signatures of



**Figure 4.7:** HB matrices of C2Q26 and C2Q36 at temperatures below the transition temperature, corresponding to the system’s ordered states.

whole-chain  $\beta$ -sheet aggregation. The full main-diagonal signatures present in the C2Q14 and C2Q16 systems are absent in C2Q26 and C2Q36. However, signatures of aggregated  $\beta$ -hairpins are still clearly visible. As in the systems discussed earlier, the  $\beta$ -sheets are aligned in an anti-parallel manner. They appear in the intra- and inter-quadrants as diagonals of descending slopes and a HB pattern of  $(i+2, i-2)$ . Additionally, both C2Q26 and C2Q36 show diagonals with ascending slope in their inter-quadrants, signifying  $\beta$ -sheets in parallel alignment. The corresponding HB pattern is  $(i+2, i+2)$ . In the C2Q26 system, four such ascending diagonals are visible, starting with HB bonds  $(1, 29)$ ,  $(14, 30)$ ,  $(27, 1)$ , and  $(28, 14)$ . In the C2Q36 system, two ascending diagonals are present, starting with HB bonds  $(8, 40)$  and  $(40, 8)$ . Here,  $(i, j)$  denotes the HB between the NH bead of residue  $i$  and the CO bead of residue  $j$ . Since these signatures of parallel  $\beta$ -sheets are only present in the inter-quadrants, they indicate inter-molecular  $\beta$ -sheet aggregation and correspond to configurations of  $\beta$ -hairpins that aggregate to form a parallel  $\beta$ -sheet between them. However, these parallel aggregated  $\beta$ -hairpins exhibit a lower probability compared to those with anti-parallel alignment. The absence of whole-chain  $\beta$ -sheet aggregation in C2Q26 and C2Q36 aligns well with the simulation results from Chen et al. [164].

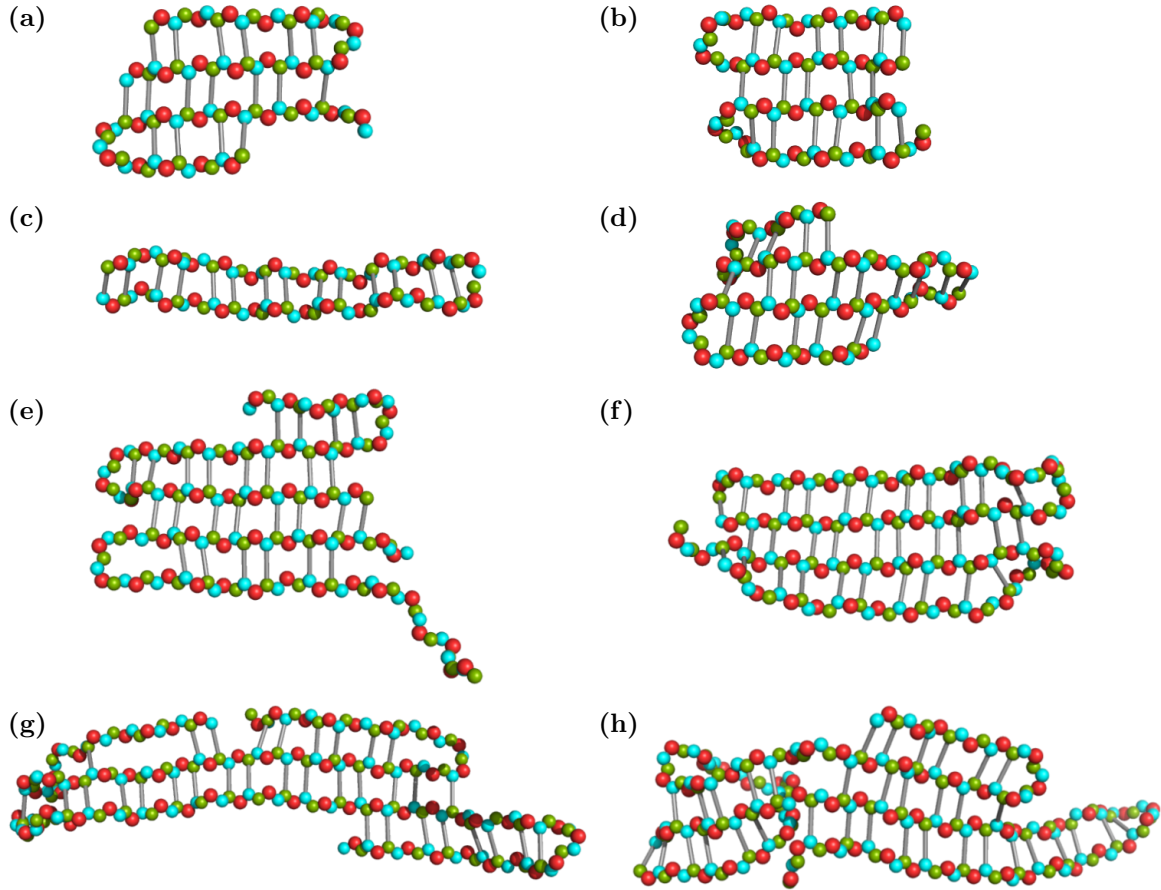
Furthermore, the HB maps of C2Q26 and C2Q36 display a greater variety of possible  $\beta$ -hairpin configurations compared to the shorter chain systems. Notably, more than one  $\beta$ -turn can occur within a single chain. This result in configurations like an S-shape with a 3-stranded  $\beta$ -hairpin, as illustrated in fig. 4.8(e). Such 3-stranded  $\beta$ -hairpins have been observed in the simulations of C1Q32 in the PRIME20 model by Marchut et al. [111]. Another example of multiple  $\beta$ -turns within a single chain is shown in fig. 4.8(g), where two  $\beta$ -hairpins form on different sides of the same chain. Aggregates of single-turn  $\beta$ -hairpins, like the one in fig. 4.8(f) are also present in systems with longer chains.

In addition to an increased number of possible  $\beta$ -turns in the folded chain, longer chains can exhibit more complex structural motifs. For instance, loops in a  $\beta$ -strand, where one chain loops over to the opposite side of the other (see fig. 4.8(h)). Despite this complexity, these configurations still align with the aggregated  $\beta$ -hairpins signatures in the HB maps.

HB maps of the ordered state of the other five PolyQ systems C2Q18, C2Q20, C2Q22, C2Q24 and C2Q28 can be found in fig. B.5 in appendix B.

### 4.2.3 Intra- vs. inter-molecular contacts

Up to this point, the analysis revealed a PT in the PolyQ systems from a random coil to a folded and aggregated state, containing  $\beta$ -sheets. However, the folding and aggregation transition occur



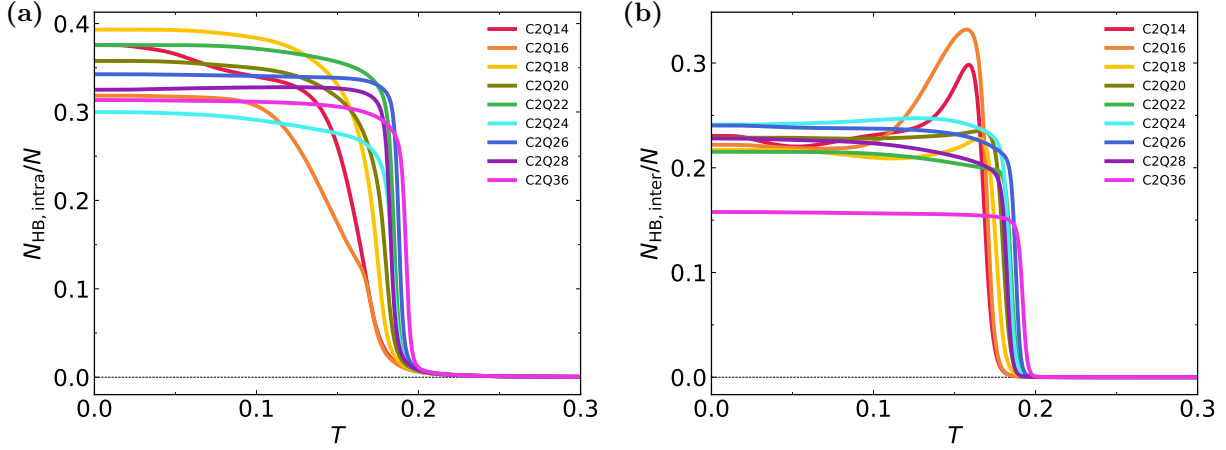
**Figure 4.8:** Configuration snapshots of folded and aggregated states for various systems. Side chain beads are hidden to improve visibility. Bead colors correspond to different backbone bead types, consistent with fig. 2.4(a). The PolyQ systems of the shown structures are: C2Q14 for (a), (b) and (c); C2Q16 for (d); C2Q26 for (e) and (f); C2Q36 for (g) and (h). The figure were taken from Lauer et al. [56].

at the same temperature. In order to distinguish the two processes, the intra- and inter-molecular contacts will be examined separately. First, the number of intra- and inter-molecular HBs will be analyzed, as HBs are the main stabilizing force of secondary structure formation. Second, the number of intra- and inter-molecular SC contacts will be investigated.

### Hydrogen bond contacts

In fig. 4.9(a) the number of intra-molecular HBs  $N_{\text{HB,intra}}$  is shown vs. the temperature  $T$ .  $N_{\text{HB,intra}}$  is normalized by the number of residues  $N$  in the systems. In the high-temperature region, all systems are in the random coil state and no intra-molecular HBs are formed. At the transition temperature  $T^*$ ,  $N_{\text{HB,intra}}$  increases rapidly, which corresponds to the collapse and folding of the single chain into  $\beta$ -hairpins, as was already shown in the HB map analysis. The inflection points of  $N_{\text{HB,intra}}(T)$  coincide well with  $T^*$ . Below  $T^*$ , the  $N_{\text{HB,intra}}$  curves stabilize at a plateau. All system sizes have a monotonous increase in their  $N_{\text{HB,intra}}$ , when going from high to low temperatures.

The  $N_{\text{HB,intra}}(T)$  curves of C2Q14 and C2Q16 exhibit additional features. At the low-temperature transition of C2Q14, at  $T_{14,\text{low}}^* = 0.060$ ,  $N_{\text{HB,intra}}$  shows a further increase. This matches the results of the HB maps analysis, where the less optimal off center  $\beta$ -hairpin signatures disappear at  $T_{14,\text{low}}^*$ , leading to an increased number of intra-molecular HBs. Furthermore, the  $N_{\text{HB,intra}}$  of C2Q16 shows a change in slope just below  $T_{16}^* = 0.174$ . This is consistent



**Figure 4.9:** Number of intra- and inter-molecular HBs vs. the temperature, normalized by the number of residues  $N$  in the systems.

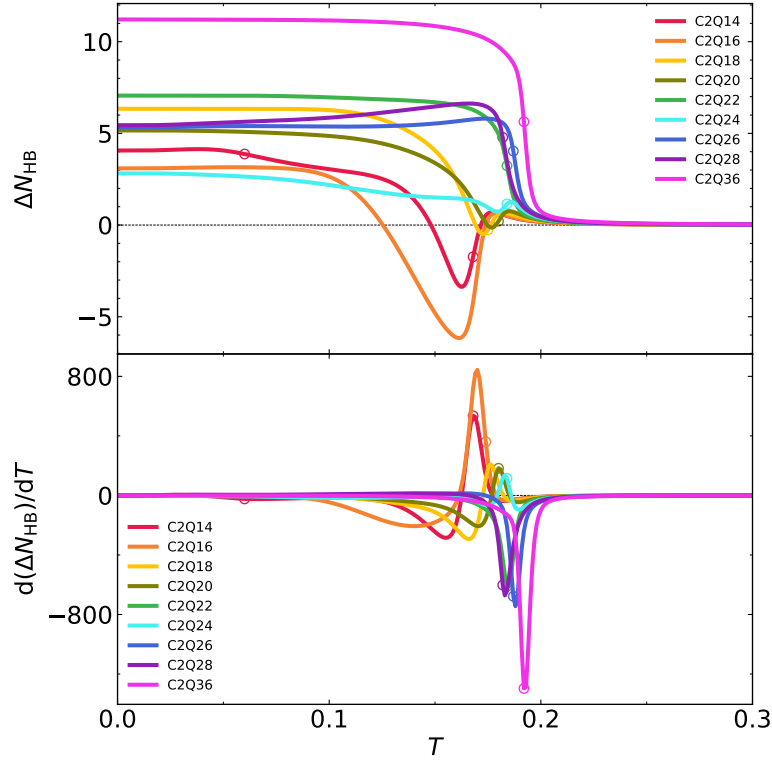
with the initial formation of the  $\beta$ -hairpins alongside whole-chain  $\beta$ -sheets. As discussed in section 4.2.2, the latter disappear when cooling further below  $T_{16}^*$  in favor of the aggregated  $\beta$ -hairpins which increase the number of intra-molecular HBs.

The number of inter-molecular HBs  $N_{\text{HB,inter}}$  is shown in fig. 4.9(b). Like in the case of intra-molecular HBs, in the high-temperature regime the chains show no inter-molecular HBs contacts as the chains are in a state of separated random coils. When cooling past  $T^*$ , the increase of  $N_{\text{HB,inter}}$  confirms the previous assignment of an aggregation transition. Below  $T^*$ , the  $N_{\text{HB,inter}}$  curves of most systems stabilize at a plateau. Again, the C2Q14 and the C2Q16 system show slightly different behavior. Both systems have a very distinct maximum in  $N_{\text{HB,inter}}$  just below their respective  $T^*$ . This maximum is followed by a decrease in  $N_{\text{HB,inter}}$  when going to lower temperatures until the curves stabilize at a plateau. The maxima in  $N_{\text{HB,inter}}$  are at a higher value than the normalized  $N_{\text{HB,inter}}$  of the other systems.

The maxima in  $N_{\text{HB,inter}}$  of C2Q14 and C2Q16 are a result of the formation of whole-chain  $\beta$ -sheet structures, as discussed in the HB map analysis. When these structures disappear at lower temperatures, the  $N_{\text{HB,inter}}$  decrease, while the  $N_{\text{HB,intra}}$  increase as  $\beta$ -hairpins are formed. The number of inter-molecular HBs stabilizes until it is in line with the percentage of inter-molecular HBs of the other systems. This percentage is around and slightly less than 25% of the total number of residues in the system. Two things need to be considered when interpreting this value: First, that for aggregated single-turn  $\beta$ -hairpins, half of both chains are facing each other to form the inter-molecular  $\beta$ -sheet. Second, in a  $\beta$ -sheet, only every second residue along the chain is involved in the HB formation. This results in the percentage of  $\sim 25\%$  of inter-molecular HBs in the system. The C2Q36 system deviates from this behavior, as it has a lower percentage of inter-molecular HBs at low temperatures. Since C2Q36 only rarely forms centered single-turn  $\beta$ -hairpins, the percentage of inter-molecular HBs remains below 25%.

In order to compare the intra- and inter-molecular HBs formation, the difference  $\Delta N_{\text{HB}} = N_{\text{HB,intra}} - N_{\text{HB,inter}}$  is shown in the top plot of fig. 4.10. Positive values correspond to a dominance of inter-molecular HBs, whereas negative values indicate a dominance of intra-molecular HBs. The PT temperatures are marked by circles. In the high-temperature regime,  $\Delta N_{\text{HB}}$  is positive, since the chains are separated and only intra-molecular HBs occasionally form. When cooling towards  $T^*$ ,  $\Delta N_{\text{HB}}$  shows different behavior for the different chain lengths. The systems can be divided into two groups.

The first group consists of shorter chains with  $N_{\text{res}} \in (14, 16, 18, 20, 24)$ . The  $\Delta N_{\text{HB}}$  graphs of these systems show a positive slope at  $T^*$ . Additionally, they display a minimum in  $\Delta N_{\text{HB}}$  at a temperature below  $T^*$ , after which  $\Delta N_{\text{HB}}$  increases once again. This becomes more evident, when looking at the derivative  $d(\Delta N_{\text{HB}})/dT$ , shown in the bottom plot of fig. 4.10. Here, the



**Figure 4.10:** The top plot shows the difference between the number of intra- and inter-molecular HBs as a function of temperature. The bottom plot shows the derivative of the top plot. The circles mark the transition temperatures  $T^*$  of the systems.

graphs of systems from the first group have a positive value at  $T^*$ . The positive slope suggests, that at  $T^*$ , the formation of inter-molecular HBs, i.e. the aggregation of the two chains, is the dominant process, driving the PT. Below  $T^*$ , the derivatives of the systems from the first group are negative, indicating a dominance of intra-molecular HBs formation. This corresponds to the formation of aggregated  $\beta$ -hairpins discussed in the HB map analysis.

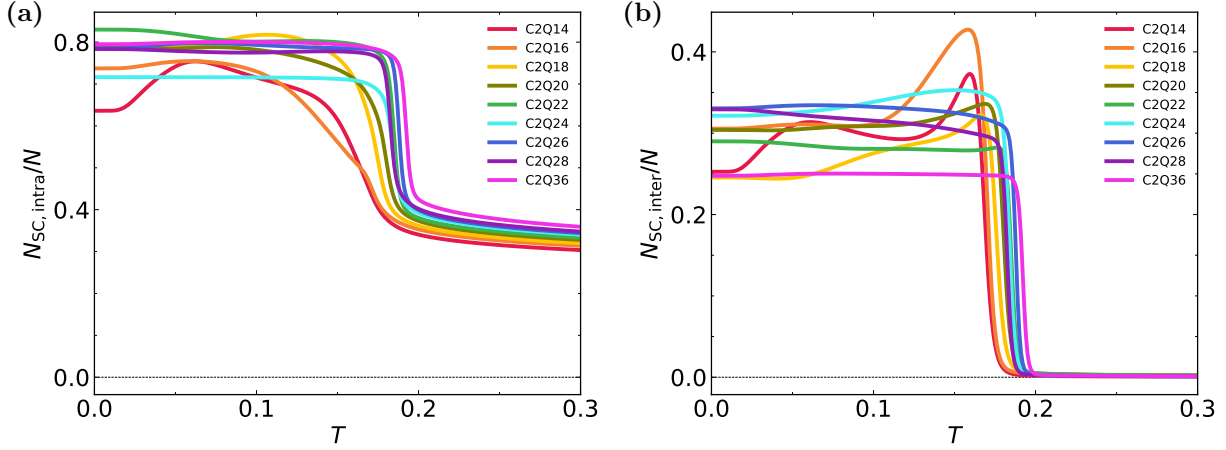
The second group consists of longer chains with  $N_{\text{res}} \in (22, 26, 28, 36)$ . For these systems,  $\Delta N_{\text{HB}}$  is positive over the entire  $T$  range. The  $\Delta N_{\text{HB}}$  graphs exhibit a negative slope at  $T^*$ , which translates to negative values of their derivatives. This indicates, that in the longer chain systems, the formation of intra-molecular HBs, i.e. the folding of the single chains, is the process dominating the PT.

It is worth noting, that the assignment of the chain lengths to the groups is non-consecutive. The C2Q22 system is part of the long-chain group, while the C2Q24 system is part of the short-chain group. This suggests, that there is a transition region between the two groups, in which a systems group affiliation can not be assigned without some ambiguity.

### Side chain contacts

Next to HB contacts, SC contacts are another indicator for folding and aggregation. Equivalent to the HB contacts, the number of intra- and inter-molecular contacts, normalized by the total number of residues  $N$ , is investigated. It is important to note that, contrary to the HBs, SCs can establish contacts with multiple partner beads at the same time. The upper limit of the number of contact a single SC bead can form is only defined by spacial exclusion effects of the hard spheres. Thus, the normalization by the total number of side chains does not directly translate to a percentage of total possible contacts.

In fig. 4.11(a), the number of intra-molecular SC contacts  $N_{\text{SC,intra}}$  is shown. In the high-temperature regime,  $N_{\text{SC,intra}}$  is non-zero. The chains are in a random coil state in which

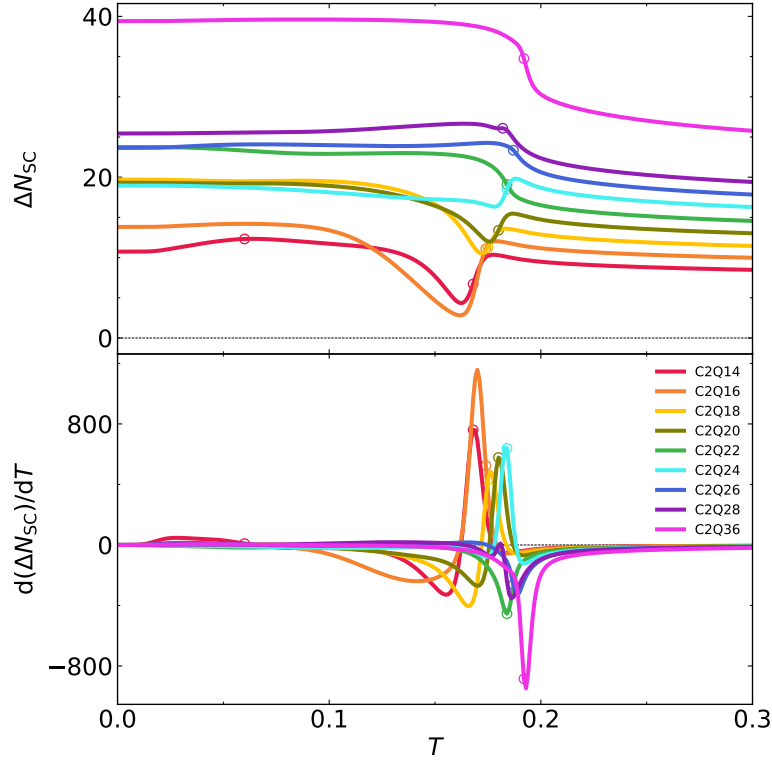


**Figure 4.11:** Number of intra- and inter-molecular side chain contacts vs. the temperature, normalized by the number of residues  $N$  in the systems.

configurations with some degree of SC contacts are the most probable, as was already discussed in section 4.1. At  $T^*$ ,  $N_{SC,intra}$  increases rapidly, indicating a collapse of the single chains. However, the slope of this increase is different depending on the chain length. Longer chains exhibit a steep increase, while short chains show a slower increase. Below  $T^*$ , the  $N_{SC,intra}$  curves stabilize at a plateau. Longer chains reach this plateau at a higher  $T$  than shorter chains. The C2Q14 and the C2Q16 system differ from this plateau behavior, as they show a maximum in  $N_{SC,intra}$  at low temperatures. For the C2Q14 system, this maximum coincides with the low-temperature transition  $T_{14,low}^* = 0.060$ . Below this maximum,  $N_{SC,intra}$  decreases rapidly in the C2Q14 system. This suggests that the restructuring of the chains to centered  $\beta$ -hairpins, which increased the intra-molecular HBs, leads to a decrease in intra-molecular SC contacts. The restructuring also decreases the inter-molecular SC contacts, as will be discussed below.

The number of inter-molecular SC contacts  $N_{SC,inter}$  is shown in fig. 4.11(b). It shows similar behavior to  $N_{HB,inter}$  in fig. 4.9(b). In the high-temperature regime,  $N_{SC,inter}$  is near zero, as the chains are separated. At  $T^*$ , the chains aggregate and  $N_{SC,inter}$  increases. The short chain systems with  $N_{res} \in (14, 16, 18, 20, 24)$  exhibit a maximum in  $N_{SC,inter}$  just below  $T^*$ . The maximum peak is most pronounced for C2Q14 and C2Q16, which affirms the formation of whole-chain  $\beta$ -sheets in these systems. Below the PT temperature,  $N_{SC,inter}$  of most chain systems remains at a high plateau. However,  $N_{SC,inter}$  of systems with  $N_{res} \in (14, 16, 18)$  decreases significantly over a longer  $T$  range, below  $T^*$ . The C2Q14 system shows another local maximum in  $N_{SC,inter}$  around  $T_{14,low}^* = 0.060$ , followed by a steep decrease at lower temperatures. The simultaneous decrease in intra- and inter-molecular SC contacts indicates, that the configurations which become the dominant folded structure in C2Q14 below  $T_{14,low}^*$ , have fewer SC contacts while increasing the number of HBs. This is consistent with the restructuring of the chains to centered  $\beta$ -hairpins, which was observed in the HB map analysis.

Finally, the difference  $\Delta N_{SC} = N_{SC,intra} - N_{SC,inter}$  is displayed in the top plot of fig. 4.12. Across the entire temperature range, all systems show a predominance of intra-molecular SC contacts. The PT temperatures are marked by circles. Similar to the HBs, the systems can be categorized into two groups. Systems with  $N_{res} \in (14, 16, 18, 20, 24)$  exhibit a positive slope of  $\Delta N_{SC}$  at  $T^*$ , whereas systems with  $N_{res} \in (22, 26, 28, 36)$  show a negative slope. This supports the previous identification of the driving process of the folding and aggregation transition. Specifically, in short chain systems, the transition is dominated by the formation of inter-molecular HBs which corresponds to aggregation. Conversely, in long chain systems, the transition is dominated by the formation of intra-molecular HBs, which corresponds to folding of single chains.



**Figure 4.12:** The top plot shows the difference between the number of intra- and inter-molecular side chain contacts as a function of temperature. The bottom plot shows the derivative of the top plot. The circles mark the transition temperatures  $T^*$  of the systems.

### 4.3 Temperature mapping

In the previous sections, the PT in PolyQ dimer systems has been analyzed. The PT temperature  $T^*$  of the folding and aggregation transition has been determined from the canonical heat capacity analysis and structures in the folded and aggregated state have been identified. However, for a meaningful interpretation of the results it is necessary to convert the model units for  $T$  and  $E$  into physical units  $T'$  and  $E'$ . As previously described in section 2.3.1, the temperature conversion follows  $T' = \epsilon_{\text{HB}} T / k_B$ . Ergo, it requires determining the value of  $\epsilon_{\text{HB}}$ , which is the energy scale of the HB interactions. As the PRIME20 model is an implicit-solvent model, the  $\epsilon_{\text{HB}}$  also takes solvent-peptide interactions, as well as HB interactions of SCs into account.

The temperature conversion for PolyQ systems in the PRIME20 model has been performed previously by Böker et al. [53]. The resulting temperature conversion formula is

$$T'[K] = 1650K \cdot T. \quad (4.1)$$

Following this equation, room temperature (300K) corresponds to  $T = 0.182$  in reduced PRIME20 model units. Consequently, the PTs of the PolyQ systems shown in both plots in fig. 4.3 occur around room temperature. This is reasonable, as the aggregation behavior of PolyQ is associated with amyloid disorders in the human body. Using the linear extrapolation in fig. 4.3(b), the PolyQs systems can again be divided into a short-chain and a long-chain group. The PTs of the short-chain systems with  $N_{\text{res}} \leq 21$  are below room temperature, while the PTs of the long-chain systems with  $N_{\text{res}} \geq 22$  are above room temperature. This result is consistent with the experimental findings of Kar et al. [36], which show that, at physiological temperatures, there is a critical chain length of around  $N_{\text{res}} = 24$ , above which the single PolyQ chain can act as a nucleation site for the aggregation of multiple chains.

## 4.4 Conclusions

Using the SAMC simulation method and the PRIME20 protein model, dimer systems of PolyQ chains were simulated. The chains in the system were of lengths  $N_{\text{res}} \in (14, 16, 18, 20, 22, 24, 26, 28, 36)$ . From the SAMC procedure, the microcanonical entropy  $S(U)$  was obtained, enabling a comprehensive thermodynamic analysis of the systems. Peaks in the canonical heat capacity  $C_V(T)$  revealed a PT in the systems with a transition temperature  $T^*$  that depends on  $N_{\text{res}}$ . When converting the reduced model temperature to physical temperature, the PTs were found to occur around room temperature. Specifically, short chain systems with  $N_{\text{res}} < 22$  exhibit a PT temperature below room temperature, while long chain systems with  $N_{\text{res}} \geq 22$  have a PT temperature above room temperature. That PolyQ chains are in a collapsed state at room temperatures has been suggested previously by Walters et al. [26] in their fluorescence resonance energy transfer studies.

Analysis of the tensor of gyration revealed that the single chains form elongated structures at the PT, consistent with the behavior of  $\beta$ -hairpins. By examining contact probabilities between residues via HB maps, the PT was characterized as folding and aggregation transitions. The HB maps demonstrated that short chain systems form both aggregated  $\beta$ -hairpins and whole-chain  $\beta$ -sheets, while long chain systems only form aggregated  $\beta$ -hairpins. These findings align with experimental results by Kar et al. [36] and simulations by Chen et al. [164], which also observed  $\beta$ -sheets and  $\beta$ -hairpins.

In the canonical analysis, the folding and aggregation transition appear as a single peak in the heat capacity. However, the microcanonical heat capacity reveals two distinct transitions occurring at the same temperature. Analysis of intra- and inter-molecular HB and SC contacts indicates that the folding and aggregation transition is driven by different processes in short and long chain systems. In short chain systems with  $N_{\text{res}} \leq 24$ , the PT is driven by the formation of inter-molecular HB and SC contacts, suggesting that aggregation is the dominant process behind the PT. Conversely, in long chain systems with  $N_{\text{res}} > 24$ , the PT is driven by the formation of intra-molecular HB and SC contacts, indicating that single chain folding is the dominant process. This finding confirms previous experimental and simulation results that identified  $N_{\text{res}} = 24$  as the critical nucleus size for single chain aggregation nucleation of PolyQ [26, 36, 37, 164].

## Chapter 5

# Dimerization of Amyloid Beta

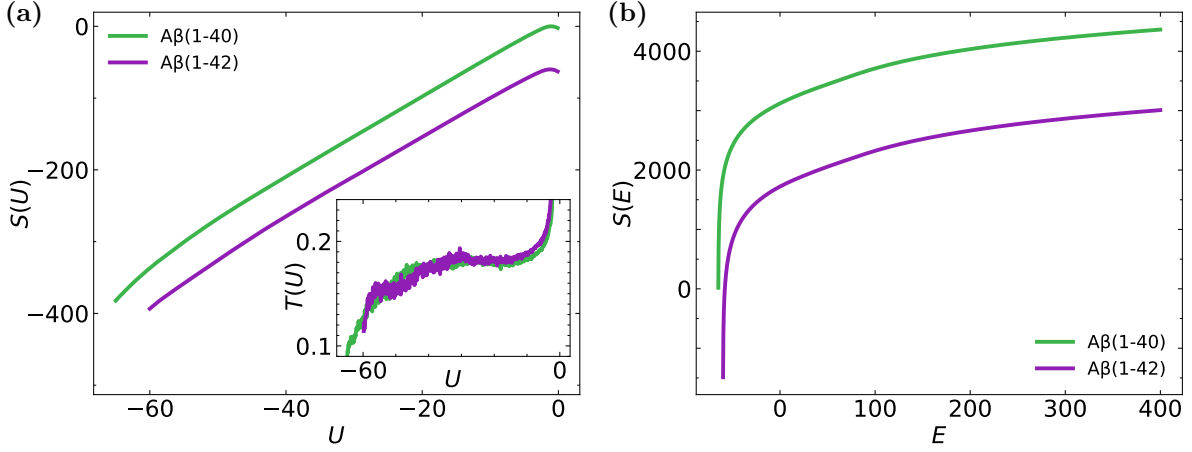
This chapter investigates the thermodynamic and structural properties of Amyloid beta ( $A\beta$ ) protein dimer systems. The  $A\beta$  protein comes in two isoforms,  $A\beta(1-40)$  and  $A\beta(1-42)$ , which differ in length. The amino acid sequence of  $A\beta(1-42)$  is presented in eq. (3.31).  $A\beta(1-40)$  shares the same sequence but lacks the final two residues, isoleucine and alanine. Both isoforms are known to play a crucial role in the pathogenesis of Alzheimer’s disease. However, the aggregation properties of the two forms differ, with  $A\beta(1-42)$  being more prone to aggregation and considered more toxic than  $A\beta(1-40)$  [23, 24]. Consequently, it is essential to understand the distinct aggregation behaviors of these two isoforms.

The conducted investigation of the  $A\beta$  systems follows a similar approach to that used for the PolyQ systems. Initially, the thermodynamic properties of the systems are investigated, including an analysis of entropy, temperature, and heat capacity. Subsequently, a structural analysis is performed, which includes an evaluation of the tensor of gyration, as well as a quantification of intra- and inter-molecular hydrogen bond and side chain contacts. Additionally, this chapter utilizes a secondary structure analysis based on the evaluation of Ramachandran plots for the  $A\beta$  systems.

### 5.1 Thermodynamic analysis

As was shown in section 3.3, the SAMC simulation method produces an approximation of the logarithm of the microcanonical density of states  $g(U)$ , which is equivalent to the microcanonical entropy  $S(U) = \ln g(U)$ . The two approximated entropies for the  $A\beta(1-40)$  and the  $A\beta(1-42)$  system are presented in fig. 5.1(a). Both exhibit a very similar slope and shape. The maximum entropy is located at high energies, though notably below zero. This suggests that the most probable configurations of the system do not correspond to the maximum energy. Instead, these configurations correspond to states where some beads are in energetically favorable contact. This phenomenon already appeared in the PolyQ systems and has already been explored in greater detail in section 4.1. Simulations for  $A\beta(1-40)$  converged at lower energies compared to  $A\beta(1-42)$ . Therefore,  $S(U)$  was approximated over the interval  $[-65.0; 0.0]$  for  $A\beta(1-40)$  and over  $[-60.0; 0.0]$  for  $A\beta(1-42)$ .

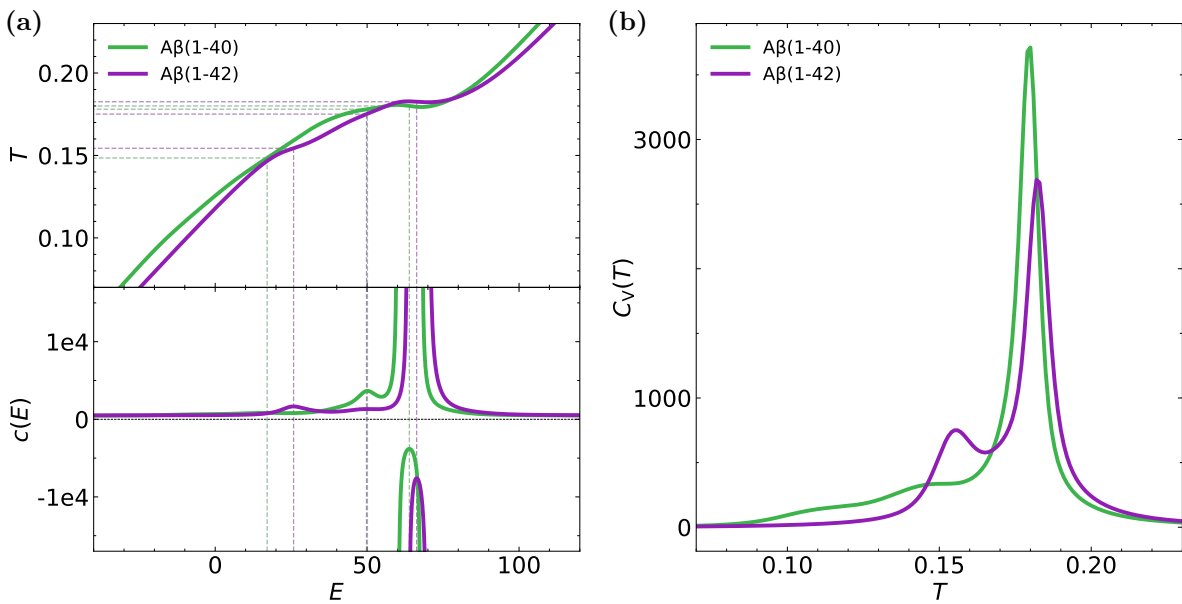
Although the  $S(U)$  curves in fig. 5.1(a) appear smooth at first glance, closer examination reveals small-scale fluctuations, which become more apparent when considering the derivative  $dS/dU = 1/T$ , or rather its inverse, the microcanonical temperature  $T(U)$ . The  $T(U)$  plot for the two  $A\beta$  systems is depicted in the inset in fig. 5.1(a). The noisy behavior in the  $T(U)$  plot reveals the subtle fluctuations in  $S(U)$ , that are not immediately evident in the  $S(U)$  curves themselves. In contrast to the PolyQ systems, where  $S(U)$  oscillates and changes the sign of its slope,  $S(U)$  for the  $A\beta$  systems increases monotonically from the minimum energy to the energy of the global maximum. Consequently, the  $T(U)$  curves do not display the wild oscillations between positive and negative values observed in the PolyQ example of the C2Q16



**Figure 5.1:** Configurational and total-energy microcanonical entropy of both A $\beta$  systems. (a) Entropy  $S(U)$  as a function of energy  $U$ . Inset: Microcanonical temperature  $T(U)$ . (b) Entropy  $S(E)$  as a function of total energy  $E$ .

system in fig. 4.2(a). However, the noisy fluctuations in  $T(U)$  of the A $\beta$  systems make further thermodynamic analysis on this representation challenging. To address this, the entropies were transformed from the configurational microcanonical ensemble to the full-energy microcanonical ensemble using the convolution method described in eq. (3.33). The resulting  $S(E)$  curves, shown in fig. 5.1(b), are smooth, enabling more straightforward thermodynamic analysis.

For an interpretation of the thermodynamic properties of the A $\beta$  systems, the temperature  $T(E)$  and heat capacity  $c(E)$  are analyzed. The temperature in the microcanonical ensemble is depicted at the top of fig. 5.2(a), with the heat capacity shown below. Changes in the curvature of  $T(E)$  indicate PTs, which translates to peaks in the heat capacity  $c(E)$ . The peak positions and the associated PT temperatures are marked by colored, dashed lines in fig. 5.2(a). Both systems exhibit a negative maximum in  $c(E)$ , which is indicative of a first-order PT. These translate to a temperature around  $T_{\text{high}}^* \approx 0.18$ . Additionally, each of the A $\beta$  systems displays two positive maxima in their  $c(E)$  curves, suggesting the presence of two



**Figure 5.2:** Plots relevant to the thermodynamic analysis of the A $\beta$  systems. (a) Microcanonical temperature  $T(E)$  and heat capacity  $c(E)$ . (b) Canonical heat capacity  $C_V(T)$ .

**Table 5.1:** Phase transition temperatures  $T^*$  of both  $A\beta$  systems as obtained from the microcanonical and canonical analysis.

System	$T_{\text{micro}}^*$			$T_{\text{can}}^*$
	1 <sup>st</sup> order	2 <sup>nd</sup> order	Gibbs-loop	
$A\beta(1-40)$	—	0.1485	—	—
	0.1800	0.1780	0.1800	0.180
$A\beta(1-42)$	—	0.1543	—	0.155
	0.1826	0.1751	0.1825	0.182

second-order PTs. All PT temperatures derived from the heat capacity analysis are summarized in table 5.1. The temperature of the higher-energy second-order maxima align with the first-order PT temperatures, suggesting that two transitions occur almost simultaneously around  $T_{\text{high}}^*$ . On the other hand, the lower-energy second-order maxima correspond to temperatures of approximately  $T_{\text{low}}^* \approx 0.15$ . Notably, the two second-order peaks in the heat capacity show different intensity between the  $A\beta$  systems. In the case of  $A\beta(1-40)$ , the high-energy second-order peak is more prominent than the low-energy second-order peak. In contrast, for  $A\beta(1-42)$ , the low-energy second-order peak is more pronounced. PT temperatures determined via Gibbs-construction in the  $T(E)$  curves are also included in table 5.1. The Gibbs-construction results are in good agreement with the PT temperatures derived from the heat capacity analysis.

In addition to the microcanonical analysis, the systems can also be examined within the canonical ensemble. The heat capacity  $C_V(T)$  is derived from the canonical partition function  $Z(T)$ , as outlined in eqs. (3.39) and (3.40). The resulting  $C_V(T)$  curves are displayed in fig. 5.2(b). Peaks in  $C_V(T)$  again signal the presence of PTs. For of  $A\beta(1-40)$ , a single, sharp peak is present at  $T_{\text{can}}^* \approx 0.18$ , which signifies a first-order PT.  $T_{\text{can}}^*$  aligns with the temperature of the first-order PT that was identified in the microcanonical analysis. In contrast, the  $C_V(T)$  curve of  $A\beta(1-42)$  shows distinct peaks: one at  $T_{\text{can}}^* \approx 0.18$  and another at  $T_{\text{can}}^* \approx 0.15$ . The high-temperature peak is sharp, characteristic of a first-order PT, while the broader low-temperature peak marks a second-order PT. Both peaks align with  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ , respectively, as identified in the microcanonical ensemble. For the  $A\beta(1-40)$  system, a second-order signature corresponding to  $T_{\text{low}}^*$  is not visible in the  $C_V(T)$  curve. However, a shoulder is present at this temperature, indicating a less pronounced second-order PT. The microcanonical analysis already suggested that the second-order PT at  $T_{\text{low}}^*$  is less pronounced in the  $A\beta(1-40)$  system compared to the  $A\beta(1-42)$  system. In  $C_V(T)$  of  $A\beta(1-40)$ , another shoulder is present at a temperature of approximately  $T \approx 0.11$ , below  $T_{\text{low}}^*$ . Even though this shoulder in  $C_V(T)$  is a very faint signature, it suggests that the  $A\beta(1-40)$  system exhibits further changes in its configurations at these low temperatures. This feature in the  $C_V(T)$  curve is not present in the  $A\beta(1-42)$  system. In the remainder of this chapter,  $T_{\text{high}}^*$  will be used to refer to a system's first-order PT temperature at  $T_{\text{high}}^* \approx 0.180$  and  $T_{\text{high}}^* \approx 0.182$ , while  $T_{\text{low}}^*$  will denote a systems second-order PT temperature at  $T_{\text{low}}^* \approx 0.149$  and  $T_{\text{low}}^* \approx 0.154$  for  $A\beta(1-40)$  and  $A\beta(1-42)$ , respectively.

The thermodynamic analysis clearly reveals significant differences in the phase behavior of  $A\beta(1-40)$  and  $A\beta(1-42)$ . The two isoforms exhibit distinct PT temperatures and heat capacity profiles. As will be shown in the structural analysis (section 5.2), the PT at  $T_{\text{high}}^*$  corresponds to an aggregation and folding transition in the  $A\beta$  systems. Thus, the differences in the PT temperatures and the heat capacity profiles of the two  $A\beta$  systems suggest that their aggregation mechanisms differ. Notably, these differences arise from the addition of just two amino acids to the C-terminus of the  $A\beta(1-40)$  sequence.

### 5.1.1 Temperature mapping

To assess the physical relevance of the observed PT temperatures and the configurations in various phases of the A $\beta$  systems, it is necessary to convert the model units of temperature  $T$  and energy  $E$  into physical units, denoted as  $T'$  and  $E'$ . Equivalent to the approach for the PolyQ systems, this conversion requires the determination of the energy scale of the HB interactions  $\varepsilon_{\text{HB}}$ .

For the unit conversion, the data from Han et al. [165] was chosen. The study investigated the fibril growth of A $\beta$ (1-42) using replica exchange molecular dynamics (REMD) simulations, focusing on the interactions between a full-length A $\beta$ (1-42) monomer and a preformed A $\beta$  fibril. The fibril was constructed from two A $\beta$  fragments, consisting of residues Leu17-Ala42, which were kept in a fixed structure. In their study, Han et al. reported a docking temperature  $T_{\text{dock}}^* = 320\text{K}$ , derived from a peak in their specific heat curves. This docking temperature is taken as the equivalent to the aggregation temperature  $T_{\text{high}}^* = 0.182$  of the A $\beta$ (1-42) system, that was derived in the previous section and included in table 5.1. Based on this comparison, the temperature conversion factor ( $\varepsilon_{\text{HB}}/k_B$ ) is determined to be 1760K. The conversion rule for the temperature is then given by:

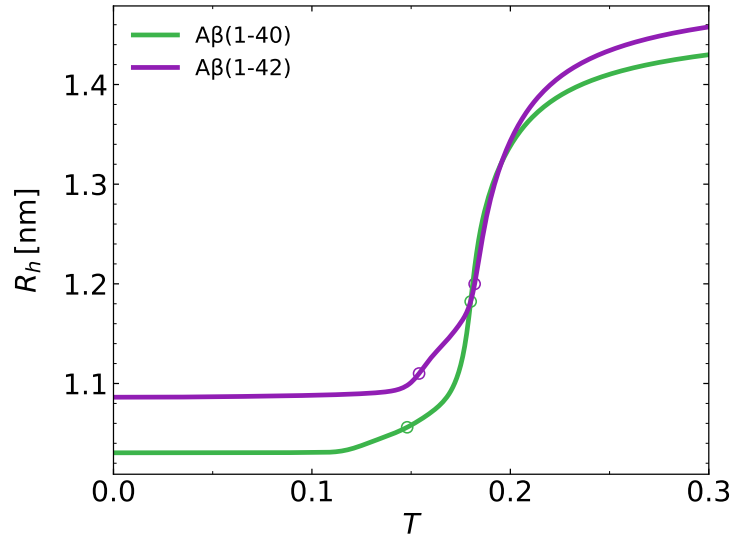
$$T' = T \cdot 1760\text{K}. \quad (5.1)$$

Using this equation, room temperature ( $T' = 300\text{K}$ ) can be expressed in the reduced units of the PRIME20 model as  $T = 0.170$ . Consequently, for both A $\beta$  systems, room temperature falls between the high and low PT temperatures,  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ . This implies that the A $\beta$  dimers exist in an aggregated state at room temperature. This is a reasonable result, considering the fact that A $\beta$  aggregates under physiological conditions in the human body.

The low-temperature second order PTs of the A $\beta$  systems translate to physical temperatures just below 0°C, namely  $T'_{\text{low}} = 261\text{K}$  for A $\beta$ (1-40), and  $T'_{\text{low}} = 272\text{K}$  for A $\beta$ (1-42). Therefore, these PTs obviously occur at non-physiological temperatures. The same is the case for the shoulder signature in the heat capacity of A $\beta$ (1-40) at  $T \approx 0.11$  ( $T' \approx 194\text{K}$ ), discussed in the previous section. At these extremely low temperatures, it is also questionable if the PRIME20 model gives results that are physically correct, since its parameters of PRIME20 were optimized with PDB structures at room temperature (see section 2.3.1). Nevertheless, these low temperature states will still be considered in the following analysis, as they can provide additional insights into the aggregation behavior of the A $\beta$  systems and their significance can not generally be dismissed. The main focus however will be on the system's aggregated and folded state that lies in the temperature range between  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ .

Other studies have been considered for the temperature conversion, but were ultimately not used due to their various shortcomings to this purpose. Takeda et al. [166] also performed REMD simulations to study the deposition of A $\beta$  monomers to preformed amyloid fibrils. However, the authors used the truncated sequence A $\beta$ (10-40) not only for the preformed fibril, but also for the monomers. Their results were therefore deemed less suited for a comparison to the present study, when compared to the results from Han et al. However, it is worth noting, that Takeda et al. reported a docking temperature  $T_{\text{dock}}^* = 360\text{K}$ , which is higher than the value reported by Han et al. When using this temperature for the conversion, the physiological relevance of the low-temperature states of the A $\beta$  systems increases, because room temperature would be at  $T = 0.150$  model temperature, which is closer to  $T_{\text{low}}^*$  of both A $\beta$  systems. Even though the study by Takeda et al. was not used for the temperature conversion, it encourages including the low-temperature states of the A $\beta$  systems in the investigation.

No doubt, the optimal way to achieve temperature mapping would be to compare the simulation results to experimental studies. However, the search for suitable experimental data was challenging due to the small size of the systems that had to be considered (monomer or dimers). Fibril systems involving a large number of A $\beta$  chains are more commonly studied. An experimental method suitable for comparison would be dynamic or static light scattering experiments



**Figure 5.3:** Hydrodynamic radius  $R_h$  of the A $\beta$  systems as a function of temperature. The circles indicate the values of  $R_h$  at the PT temperatures.

on monomeric A $\beta$  systems, which provide hydrodynamic radii  $R_h$  and radii of gyration  $R_g$ , respectively. Both of these radii can also be calculated from the simulation data: the hydrodynamic radius  $R_h$  can be calculated according to eq. (3.48) and is shown for both A $\beta$  systems in fig. 5.3, while the square of the gyration radius  $R_g^2$  is presented in the top plot of fig. 5.4(b)<sup>(4)</sup>. Both are plotted vs. the temperature  $T$ . By comparison with experimental results for  $R_h$  or  $R_g$ , the model temperature corresponding to the temperature of the experiment could potentially be determined.

Zhang-Haagen et al. [167] employed small angle neutron scattering and dynamic light scattering experiments to measure  $R_g$  and  $R_h$  for monomeric A $\beta$ (1-40) and A $\beta$ (1-42) peptides. They reported the following values for  $R_g$  and  $R_h$  at room temperature:

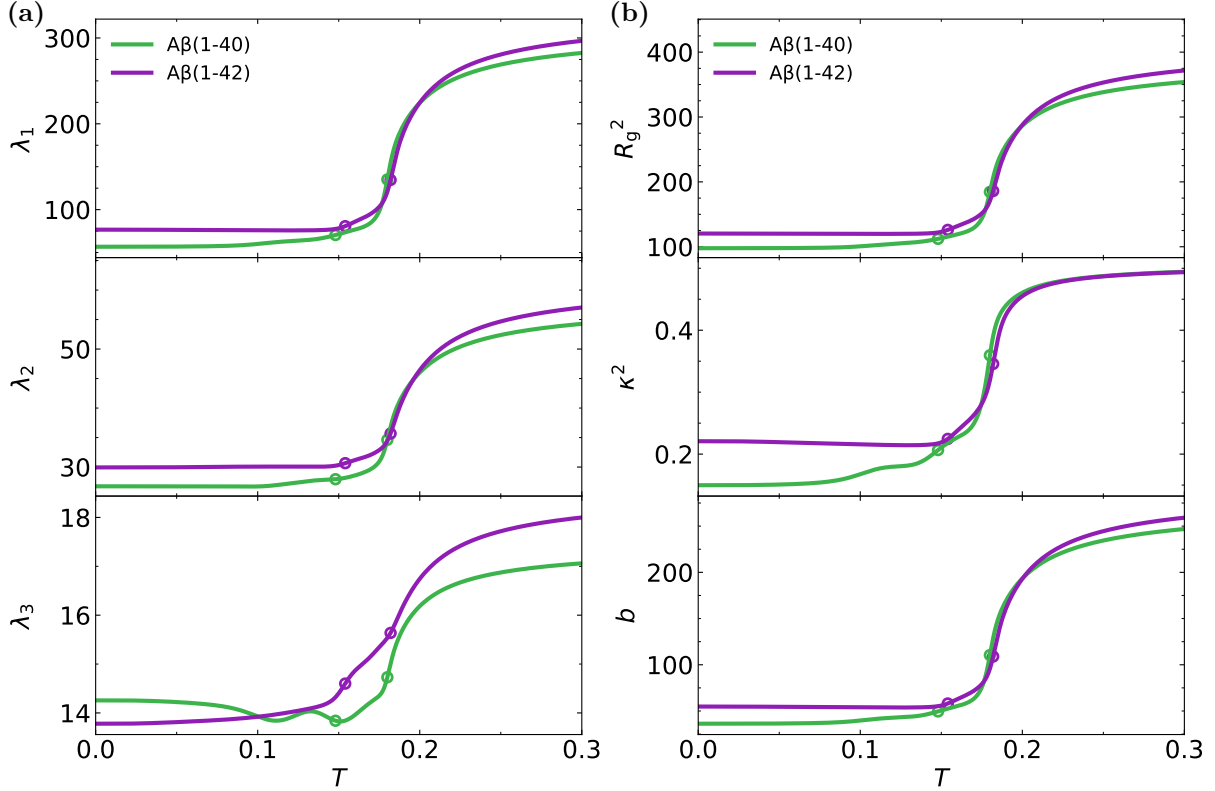
	$R_g$	$R_h$
A $\beta$ (1-40)	$1.0 \pm 0.1\text{nm}$	$1.8 \pm 0.3\text{nm}$
A $\beta$ (1-42)	$1.6 \pm 0.1\text{nm}$	$3.2 \pm 0.4\text{nm}$

The reported values for both  $R_g$  and  $R_h$  for A $\beta$ (1-42) are nearly double those of A $\beta$ (1-40), which is highly unlikely given that A $\beta$ (1-42) contains only two additional residues. This discrepancy raises doubts about the accuracy of these results. Consequently, the data from Zhang-Haagen et al. were considered unsuitable for temperature mapping in the PRIME20 model.

In another study, Nag et al. [168] used dynamic light scattering experiments to determine  $R_h$  for monomeric A $\beta$ (1-40) and A $\beta$ (1-42). For both A $\beta$  systems, they reported  $R_h = 0.9 \pm 0.1\text{nm}$ . However, these results are also not applicable for temperature mapping to the PRIME20 temperatures because they fall outside the range of  $R_h$  values obtained from the simulations (see fig. 5.3). Even with the measurement errors taken into account, the results from Nag et al. only come closest to  $R_h$  at very low temperatures, which is in the regime of near constant  $R_h(T)$  and makes unique temperature mapping impossible.

Therefore, the most suitable data source for the purpose of converting model to physical temperatures is provided by Han et al. By applying the derived conversion rule, the hydrodynamic radius from the simulation results at room temperature is calculated as  $R_h(T = 0.170) = 1.09\text{nm}$  for A $\beta$ (1-40) and  $R_h(T = 0.170) = 1.15\text{nm}$  for A $\beta$ (1-42). These values fall within the range or

<sup>(4)</sup>Attention needs to be paid to the different units used in the plots.  $R_h$  is given in nm, while  $R_g^2$  is given in  $\text{\AA}^2$ .



**Figure 5.4:** Quantities derived from the tensor of gyration as a function of temperature for both A $\beta$  systems. (a) Eigenvalues  $\lambda_i$  of the tensor of gyration  $\mathbf{T}_g$ . (b) Radius of gyration  $R_g$ , relative shape anisotropy  $\kappa^2$  and asphericity  $\Delta$ . Units:  $[\lambda_i] = \text{\AA}$ ,  $[R_g] = \text{\AA}^2$ ,  $[\kappa^2] = 1$ ,  $[b] = 1$ . Units:  $[\lambda_i] = \text{\AA}$ ,  $[R_g] = \text{\AA}^2$ ,  $[\kappa^2] = 1$ ,  $[b] = 1$ .

reported experimental values from Zhang-Haagen et al. and Han et al., supporting the validity of the conversion rule and its reasonable reproduction of experimental results.

With the established correspondence between simulation and physical temperatures, the following sections will delve deeper into the structural properties of the A $\beta$  systems to gain further insights into the differences in the aggregation behavior between A $\beta$ (1-40) and A $\beta$ (1-42).

## 5.2 Structural analysis

### 5.2.1 Tensor of gyration

First, the tensor of gyration  $\mathbf{T}_g$  is analyzed to investigate the structural properties of the A $\beta$  systems. The eigenvalues of the tensor of gyration are plotted in fig. 5.4(a). As the temperature decreases below  $T_{\text{high}}^*$ , all eigenvalues exhibit a sharp decline, indicating a collapse of the polymer chain. Above and below the transition region, the eigenvalues of A $\beta$ (1-40) are consistently lower than those of A $\beta$ (1-42) at low temperatures, which intuitively seems reasonable as A $\beta$ (1-40) is the shorter chain. At temperatures around  $T_{\text{high}}^*$  the eigenvalues  $\lambda_1$  and  $\lambda_2$  of A $\beta$ (1-40) are greater than those of A $\beta$ (1-42). This behavior can be attributed to the different transition temperatures of the first-order PT of the two A $\beta$  systems (see table 5.1). The third eigenvalue,  $\lambda_3$ , shows distinct behaviors between the two isoforms. For A $\beta$ (1-42),  $\lambda_3$  decreases monotonically as temperature decreases. Furthermore, the decrease in  $\lambda_3$  is a 2-step process, where the slope of  $\lambda_3$  flattens between  $T_{\text{low}}^*$  and  $T_{\text{high}}^*$  before steepening again below  $T_{\text{low}}^*$  and finally reaching a plateau at low  $T$ . In contrast, the behavior of A $\beta$ (1-40) is more complex:  $\lambda_3$  decreases sharply

around  $T_{\text{high}}^*$ , reaching a minimum at  $T_{\text{low}}^*$ . Subsequently,  $\lambda_3$  increases, suggesting a restructuring of the polymer chain, before decreasing again to a second minimum situated at around  $T = 0.11$ . At lowest temperatures  $\lambda_3$  stabilizes in a plateau. Notably, in this low temperature regime,  $\lambda_3$  for  $A\beta(1-40)$  is higher than that of  $A\beta(1-42)$ . The behavior of  $\lambda_3$  of  $A\beta(1-40)$  over the entire temperature range suggests a multi-step folding process. A complexity that is different to the  $A\beta(1-42)$  system.

The squared radius of gyration  $R_g^2$ , the relative shape anisotropy  $\kappa^2$  and the asphericity parameter  $b$  are quantities that can be derived from the tensor of gyration (see eqs. (3.45) to (3.47)). They are shown in fig. 5.4(b). The  $R_g^2$  graphs of both systems decrease sharply when cooling below  $T_{\text{high}}^*$ , indicating again a collapse of the chains. In both the unfolded and the folded state, the  $R_g^2$  values of  $A\beta(1-40)$  are lower than those of  $A\beta(1-42)$ . In the transition region around  $T_{\text{high}}^*$  this statement is violated, because  $T_{\text{high}}^*$  of  $A\beta(1-40)$  is lower than that of  $A\beta(1-42)$ . The shape anisotropy  $\kappa^2$  of both systems also decreases when cooling below  $T_{\text{high}}^*$ , indicating a transition from a rod-like to a more spherical shape. In the low energy regime,  $\kappa_{40}^2$  of the  $A\beta(1-40)$  system is lower than  $\kappa_{42}^2$  of the  $A\beta(1-42)$  system. However, additional distinctions between the two  $A\beta$  systems can be made. While  $\kappa_{42}^2$  exhibits a single drop region at  $T_{\text{high}}^*$  that ends around  $T_{\text{low}}^*$ , the behavior of  $\kappa_{40}^2$  is more complex. 3 distinct regions, where  $\kappa_{40}^2$  decreases rapidly can be identified. The first region is around  $T_{\text{high}}^*$ , the second region is around  $T_{\text{low}}^*$ , and the third region is at low temperatures around  $T \approx 0.10$ . In between these decrease regions, the slope of  $\kappa_{40}^2$  flattens. The positions of the inflection points of  $\kappa_{40}^2$  coincide reasonably well with the minima of  $\lambda_3$  in fig. 5.4(a). In their asphericity parameter  $b$ , both  $A\beta$  systems exhibit similar behavior.  $b$  decreases when cooling below  $T_{\text{high}}^*$ , indicating a transition from a stretched and expanded to a more compact and spherical shape. Above and below the transition temperature region,  $b$  of  $A\beta(1-40)$  is lower than that of  $A\beta(1-42)$ . The behavior of  $b$  is similar to that of  $R_g^2$ , with a sharp decrease in the transition region and a plateau at low temperatures.

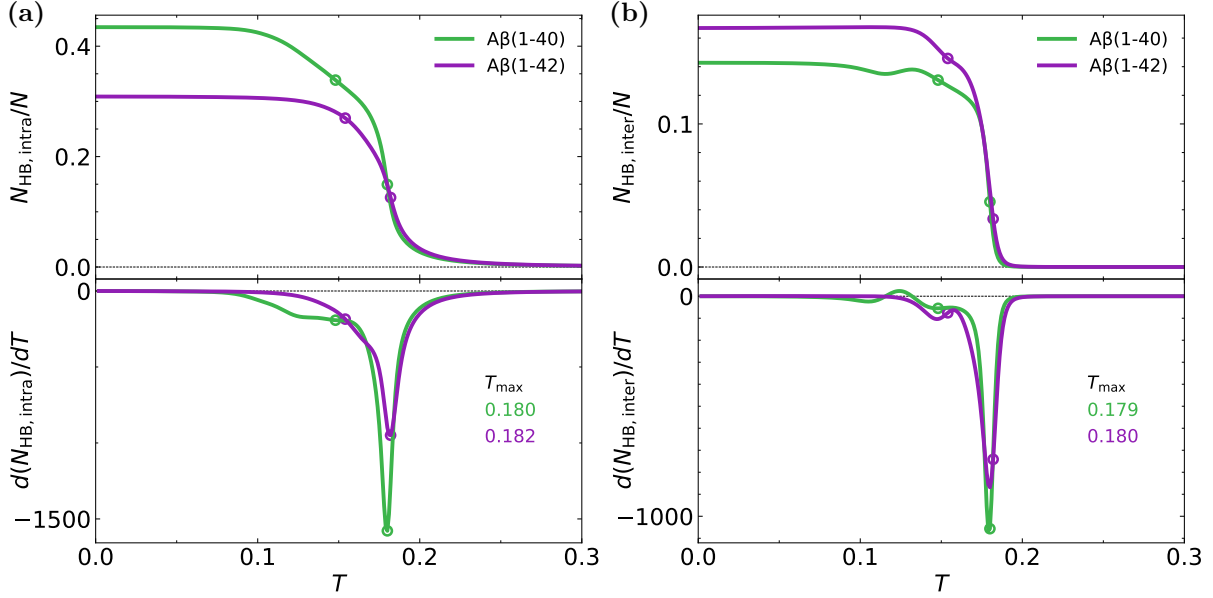
The analysis of the tensor of gyration  $\mathbf{T}_g$  highlights the distinct structural characteristics of the two  $A\beta$  isoform systems. In their folded states,  $A\beta(1-40)$  chains are more compact and spherical compared to the more elongated  $A\beta(1-42)$  chains. Additionally, the folding process of  $A\beta(1-40)$  seems to be more complex than that of  $A\beta(1-42)$ , as reflected in the behavior of the third eigenvalue  $\lambda_3$  and the asphericity parameter  $\kappa^2$ . The analysis provides valuable insights into the structural properties of individual chains within the  $A\beta$  systems. The subsequent sections will delve deeper into the intra- and inter-molecular contacts of the  $A\beta$  systems to gain a deeper understanding of their folding and aggregation behavior.

### 5.2.2 Intra vs. inter-molecular contacts

Two types of beads can establish energetically favorable contacts: hydrogen bonds (HBs) can be formed between NH and CO beads of the backbone and SC contacts can be formed between side chain beads R of the amino acids. Both are important for the protein's secondary structure formation. However, as outlined in section 2.3, HBs give a much larger contribution to the overall energy of the system when compared to SC contacts. Consequently, this section will first focus on the analysis of intra and inter-molecular HB contacts, followed by an analysis of intra and inter-molecular SC contacts.

#### Hydrogen bond contacts

Figure 5.5 displays the number of intra- and inter-molecular HB contacts in the  $A\beta$  systems versus the temperature, denoted as  $N_{\text{HB,intra}}(T)$  and  $N_{\text{HB,inter}}(T)$ , respectively. The number of contacts is normalized by the total number of residues in the system,  $N$ . As temperature decreases, both  $N_{\text{HB,intra}}$  and  $N_{\text{HB,inter}}$  increase significantly at  $T_{\text{high}}^*$ . The inflection points of these curves closely align with the PT temperatures identified in the thermodynamic analysis, as

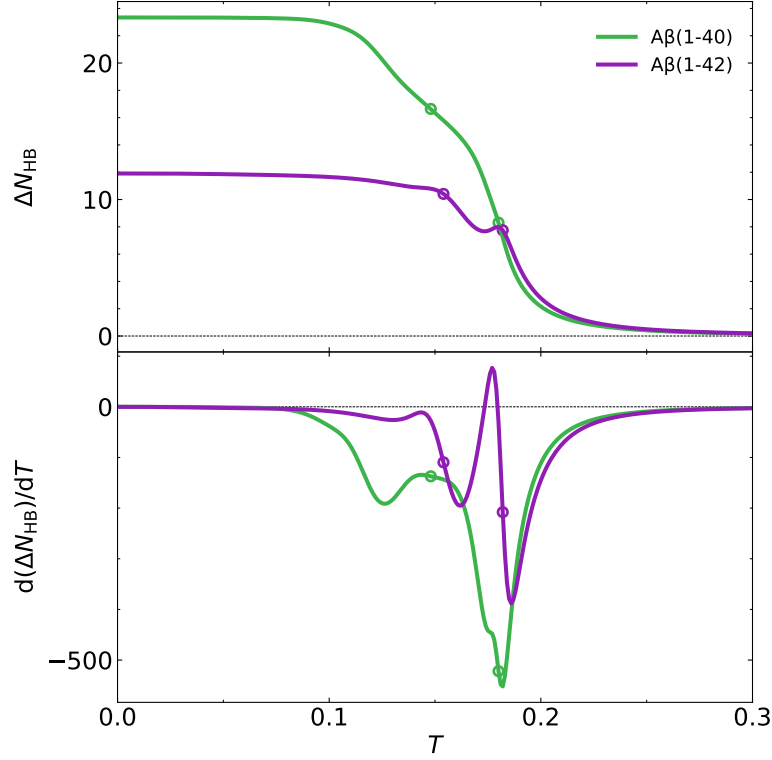


**Figure 5.5:** Number of intra- and inter-molecular HBs vs. the temperature for the A $\beta$  systems and normalized by the number of residues in the respective systems. (a) Intra-molecular HB contacts  $N_{HB,intra}(T)$ . (b) Inter-molecular HB contacts  $N_{HB,inter}(T)$ . The bottom plots show the respective derivatives.

highlighted by the derivative plots shown at the bottom of fig. 5.5. The values of  $N_{HB,intra}$  and  $N_{HB,inter}$  at the PT temperatures  $T_{high}^*$  and  $T_{low}^*$  are marked by circles. A good agreement with peaks in  $C_V(T) = \partial \langle E \rangle / \partial T$  is to be expected, since  $N_{HB,intra}$  and  $N_{HB,inter}$  are proportional to the average energy contributed by HBs in the systems. The simultaneous increase of both intra and inter-molecular contacts at  $T_{high}^*$  suggests that this transition represents both a folding and aggregation process.

The  $N_{HB,intra}$  plot in fig. 5.5(a) offers valuable insights into the collapse and folding processes of individual chains within the systems. Notably, the behavior of the two A $\beta$  systems diverges below  $T_{high}^*$ . The slope of  $N_{HB,intra}$  of A $\beta$ (1-40) between  $T_{high}^*$  and  $T_{low}^*$  decreases, becoming less steep - a change which is more clearly observed in the derivative  $d(N_{HB,intra})/dT$ . Following the initial negative peak at  $T_{high}^*$ ,  $d(N_{HB,intra})/dT$  for A $\beta$ (1-40) increases until it reaches a plateau, maintaining a constant negative value before increasing further below  $T \approx 0.12$  and reaching zero around  $T \approx 0.10$ . In contrast, A $\beta$ (1-42) exhibits a more gradual change in the slope of  $N_{HB,intra}$ . After the initial large change and a corresponding negative peak in  $d(N_{HB,intra})/dT$  at  $T_{high}^*$ , the derivative continues to decrease steadily reaching zero just below  $T_{low}^*$  and at a higher temperature than in the A $\beta$ (1-40) system. The distinct behavior of  $N_{HB,intra}$  in A $\beta$ (1-40) indicates a multi-step folding process, characterized by two distinct regimes of varying slopes around  $T_{high}^*$  and  $T_{low}^*$ . In contrast, the folding process of A $\beta$ (1-42) is more continuous. The slope of  $N_{HB,intra}$  changes gradually below  $T_{high}^*$  and through  $T_{low}^*$ .

Another key difference between the two A $\beta$  systems is the value that  $\mathcal{N}_{HB,intra} = N_{HB,intra}/N$  assumes at temperatures below  $T_{high}^*$ . For both systems,  $\mathcal{N}_{HB,intra}$  stabilizes at temperatures below  $T = 0.10$ . However, A $\beta$ (1-40) consistently exhibits a higher  $\mathcal{N}_{HB,intra}$  than A $\beta$ (1-42) across all temperatures below  $T_{high}^*$ . This indicates that in the folded state, A $\beta$ (1-40) has a greater percentage of its available HB contacts engaged in intra-molecular HB interactions compared to A $\beta$ (1-42). This contrasts with the behavior of the inter-molecular HB contacts  $\mathcal{N}_{HB,inter}$  at low temperatures. As evident in fig. 5.5(b),  $\mathcal{N}_{HB,inter}$  is lower for A $\beta$ (1-40) than for A $\beta$ (1-42) at temperatures below  $T_{high}^*$ . Consequently, the HB partners of A $\beta$ (1-42) are more involved in inter-molecular HB contacts compared to those in A $\beta$ (1-40). Since  $\mathcal{N}_{HB,intra}$  and  $\mathcal{N}_{HB,inter}$  are associated with the folding and aggregation process respectively, this difference in



**Figure 5.6:** Difference between intra and inter-molecular hydrogen bond contacts as a function of temperature. The top plot shows the difference  $\Delta N_{\text{HB}}(T) = N_{\text{HB,intra}}(T) - N_{\text{HB,inter}}(T)$ , while the bottom plot shows the derivative of  $\Delta N_{\text{HB}}(T)$ . Circles mark the transition temperatures  $T^*$  of the  $A\beta$  systems.

the behavior of intra and inter-molecular HB contacts below  $T_{\text{high}}^*$  suggests that the two  $A\beta$  systems prioritize folding and aggregation differently during the PT and different structures are formed in the folded state.

When examining  $N_{\text{HB,inter}}$  in fig. 5.5(b), another distinction between the two  $A\beta$  systems becomes evident. As the temperature decreases, both systems initially exhibit similar behavior, with  $N_{\text{HB,inter}}$  increasing at  $T_{\text{high}}^*$ . The inflection points of these curves align well with the PT temperatures derived from the thermodynamic analysis. However, as  $T$  approaches  $T_{\text{low}}^*$ , the behavior of  $N_{\text{HB,inter}}$  diverges between the systems. For  $A\beta(1-40)$ ,  $N_{\text{HB,inter}}$  changes slope as it approaches  $T_{\text{low}}^*$ . It reaches a maximum just below  $T_{\text{low}}^*$  and then decreases to a local minimum, before increasing again to its previous value at the maximum. In contrast,  $A\beta(1-42)$  shows a different behavior with different characteristic features of  $N_{\text{HB,inter}}$ . The graph increases in two steps, as indicated by 2 minima in the derivative  $d(N_{\text{HB,inter}})/dT$  at  $T_{\text{low}}^*$  and  $T_{\text{high}}^*$ . Notably, the  $A\beta(1-42)$  system experiences another significant increase in  $N_{\text{HB,inter}}$  at the second-order PT temperature, which is absent in  $A\beta(1-40)$ . This suggests that aggregation plays a more prominent role not only in the high- but also in the low-temperature transition of the  $A\beta(1-42)$  system compared to the  $A\beta(1-40)$  system.

In order to further explore the different roles of intra and inter-molecular HB contacts in the PTs of the two  $A\beta$  systems, the difference  $\Delta N_{\text{HB}}(T) = N_{\text{HB,intra}}(T) - N_{\text{HB,inter}}(T)$  was analyzed, as shown in the top plot in fig. 5.6. The first notable feature is that across the entire energy range,  $\Delta N_{\text{HB}}$  remains positive. This indicates that in both  $A\beta$  systems intra-molecular HB contacts contribute more significantly to the system's energy than inter-molecular HB contacts. At high temperatures,  $\Delta N_{\text{HB}}$  is near zero. As the temperature decreases,  $\Delta N_{\text{HB}}$  increases, indicating that the number of intra-molecular HB contacts increases more rapidly than the number of inter-molecular HB contacts. The inflection points of  $\Delta N_{\text{HB}}$  also align with  $T_{\text{high}}^*$ . Below  $T_{\text{high}}^*$ ,

the shapes of the  $\Delta N_{\text{HB}}$  curves reveal significant differences between the two systems.

In  $A\beta(1-40)$ ,  $\Delta N_{\text{HB}}$  increases monotonically as temperature decreases. Below  $T_{\text{high}}^*$ , the curve flattens before steepening again just below  $T_{\text{low}}^*$  and then reaches its maximum around  $T \approx 0.10$ , where it remains constant. This behavior, characterized by two distinct regions of maximum increase of intra-molecular HBs is clearly visible in the derivative  $d(\Delta N_{\text{HB}}(T))/dT$ . Here, two minima appear at  $T_{\text{high}}^*$  and just below  $T_{\text{low}}^*$ . At both PT temperatures, the increase of  $N_{\text{HB, intra}}$  dominates over the increase in  $N_{\text{HB, inter}}$ .

In contrast,  $A\beta(1-42)$  exhibits different behavior at the PTs temperatures. As the system approaches  $T_{\text{high}}^*$  from higher temperatures,  $\Delta N_{\text{HB}}$  initially increases. However, just below  $T_{\text{high}}^*$  the graph reaches a maximum, below which it decreases, followed by a second increase in  $\Delta N_{\text{HB}}$  at around  $T_{\text{low}}^*$ . This looped behavior is also evident in the derivative  $d(\Delta N_{\text{HB}}(T))/dT$ , where a minimum and maximum are observed between  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ . This emphasizes the importance of inter-molecular HB formation at the high-temperature transition of the  $A\beta(1-42)$  system. The subsequent increase in  $\Delta N_{\text{HB}}$  at  $T_{\text{low}}^*$  suggests that intra-molecular HB contacts become more significant at the low-temperature transition.

The analysis of intra and inter-molecular HB contacts highlights clear differences in the folding and aggregation behavior of the two  $A\beta$  systems. The  $A\beta(1-40)$  system undergoes a secondary restructuring of its HBs below the second-order PT at  $T_{\text{low}}^*$ , further favoring the formation of intra-molecular HBs. In contrast, the formation of inter-molecular HBs plays a more significant role in both the high- and low-temperature transition of the  $A\beta(1-42)$  system compared to  $A\beta(1-40)$ .

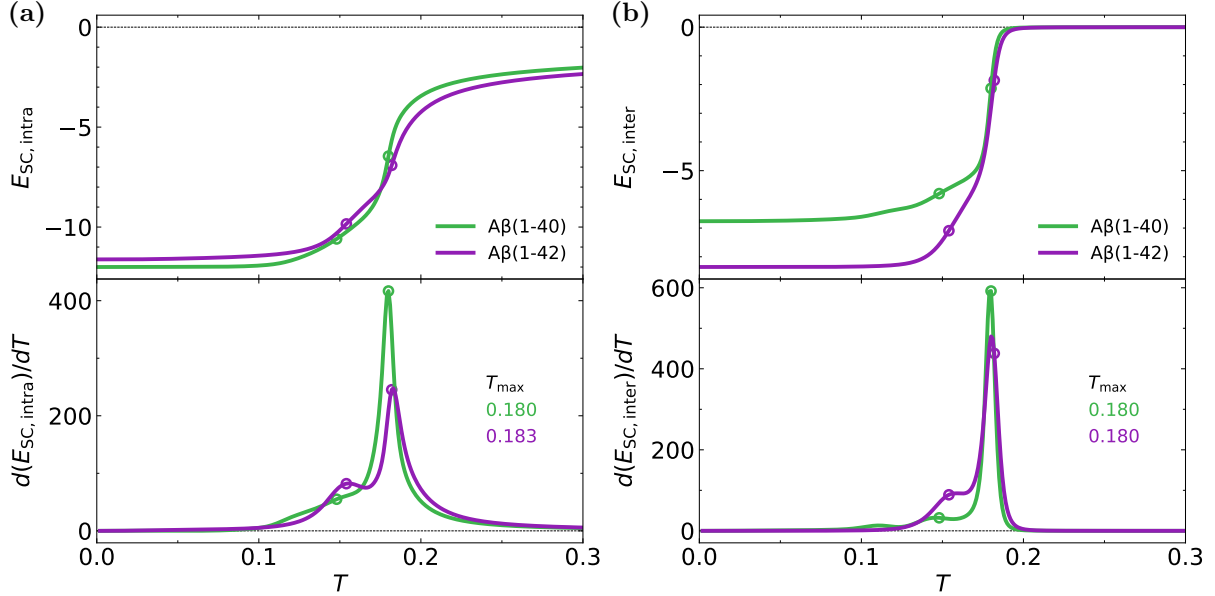
The subsequent section will discuss the analysis of intra and inter-molecular SC contacts, which, alongside the HB contacts, are contributing to the secondary structure formation of proteins.

### Side chain contacts

When looking at the intra and inter-molecular SC contacts in the  $A\beta$  systems, an important difference to the HB contacts has to be considered. While established HBs always give an energy contribution of  $E_{\text{HB}} = -1$ , the energy contribution of SC contacts  $E_{\text{SC}}$  can vary depending on the amino acids involved (see section 2.3.2 and appendix A). During the simulation runs, the total intra- and inter-molecular side chain energy was recorded. Since  $A\beta$  is a heteropolymer that contains 16 different amino acids with different  $E_{\text{SC}}$  values for different interaction energy for different pairs of amino acids, the side chain energy does not directly translate to the number of contacts. Consequently, the analysis of SC contacts will focus on the energy contributions of these contacts.

Figure 5.7 shows the intra and inter-molecular side chain energies for the  $A\beta$  systems, labeled  $E_{\text{SC, intra}}(T)$  and  $E_{\text{SC, inter}}(T)$ , respectively. As the temperature decreases, both  $E_{\text{SC, intra}}$  and  $E_{\text{SC, inter}}$  decrease significantly at  $T_{\text{high}}^*$ . In both systems, a notable change in slope of  $E_{\text{SC, intra}}$  and  $E_{\text{SC, inter}}$  occurs as the temperature approaches  $T_{\text{low}}^*$ , evidenced by peaks or shoulders in the derivative plots at the bottom of fig. 5.7. The values of  $E_{\text{SC, intra}}$  and  $E_{\text{SC, inter}}$  at the PT temperatures  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$  are highlighted by circles.

Particularly striking is the close resemblance between the derivative of  $E_{\text{intra}}$  in fig. 5.7(a) and the  $C_V(T)$  plots in fig. 5.2(b). This suggests that  $E_{\text{SC, intra}}$  closely mirrors the average energy behavior of the systems. In the  $A\beta(1-40)$  system,  $E_{\text{SC, intra}}$  decreases sharply at  $T_{\text{high}}^*$ , followed by a more gradual decrease as the temperature approaches  $T_{\text{low}}^*$ . In contrast,  $E_{\text{SC, intra}}$  in the  $A\beta(1-42)$  system exhibits two distinct temperature regimes where the energy changes, as indicated by the two peaks in the derivative. Across the entire temperature range  $E_{\text{SC, intra}}$  of both systems is quite similar, with the  $A\beta(1-40)$  system displaying slightly lower energy at low temperatures. At high temperatures, however,  $E_{\text{SC, intra}}$  in the  $A\beta(1-40)$  system is higher in the  $A\beta(1-42)$  system. Notably, in both systems  $E_{\text{SC, intra}}$  does not approach zero as temperature increases but instead stabilizes at a value corresponding to the maximum position of the entropy



**Figure 5.7:** Intra- and inter-molecular SC energies vs. the temperature for the  $A\beta$  systems. (a) Intra-molecular HB contacts  $N_{HB,intra}(T)$ . (b) Inter-molecular HB contacts  $N_{HB,inter}(T)$ . The bottom plots show the respective derivatives.

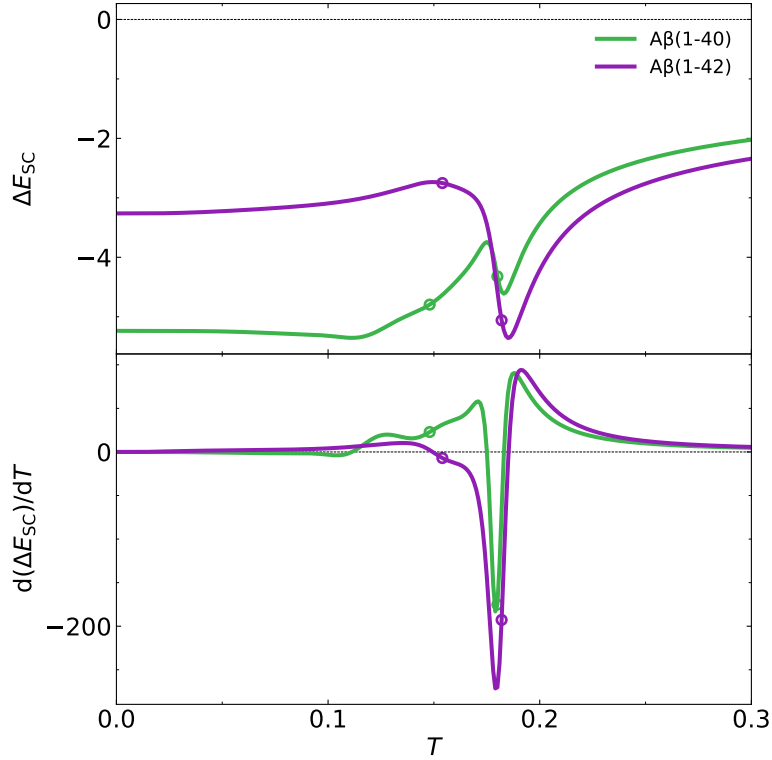
$S(U)$  in the configurational microcanonical ensemble. This supports the previous explanation that the maximum position of the  $S(U)$  being below zero indicates that entropy favors the formation of at least some SC contacts.

The behavior of  $E_{SC,inter}$ , as depicted in fig. 5.7(b), reveals more differences between the systems. At high temperatures,  $E_{SC,inter}$  of both systems remains close to zero, reflecting the separation of the chains. As the temperature approaches  $T_{high}^*$ ,  $E_{SC,inter}$  decreases, indicating the onset of chain aggregation. In both systems a temperature regime of slower decline in  $E_{SC,inter}$  is visible around  $T_{low}^*$ . At lower temperatures,  $E_{SC,inter}$  is higher in the  $A\beta(1-40)$  system compared to the  $A\beta(1-42)$  system, suggesting that  $A\beta(1-42)$  forms more SC contacts to stabilize its aggregated state.

The difference between intra and inter-molecular side chain energies  $\Delta E_{SC}(T) = E_{SC,intra}(T) - E_{SC,inter}(T)$  is shown in the top plot of fig. 5.8<sup>(5)</sup>. Its value in both  $A\beta$  systems is negative across the entire temperature range, indicating that intra-molecular SC contacts are the dominant contributors to the energy of the system. As the temperature decreases,  $\Delta E_{SC}$  decreases as well.  $\Delta E_{SC}$  reaches a minimum just above  $T_{high}^*$  below which energy contributions of inter-molecular SC contacts drive  $\Delta E_{SC}$  to increase. When approaching  $T_{low}^*$  the behavior of  $\Delta E_{SC}$  diverges between the two systems. In the  $A\beta(1-40)$  system,  $\Delta E_{SC}$  reaches a maximum after which  $E_{SC,intra}$  becomes increasingly dominant and  $\Delta E_{SC}$  decreases. It reaches a minimum around  $T \approx 0.10$  and remains at a low value in the low-temperature regime. In contrast,  $\Delta E_{SC}$  of the  $A\beta(1-42)$  system reaches its maximum approximately at around  $T_{low}^*$ , after which it roughly remains at a constant level much closer to its value above the transition than the  $A\beta(1-40)$  system. This again indicates that aggregation plays a more prominent role in both the low- and the high-temperature transition for the  $A\beta(1-42)$  system compared to the  $A\beta(1-40)$  system.

After highlighting these differences in intra- and inter-molecular HB and SC contacts, the next section will further explore the hydrogen bond networks formed in the different phases of the  $A\beta$  systems.

<sup>(5)</sup>Note that, unlike HBs, the focus here is on the energy rather than the number of formed side-chain contacts. An energetically favorable side-chain contact gives a negative contribution to  $E_{SC,intra}(T)$  and  $E_{SC,inter}(T)$ . Consequently, when interpreting  $\Delta E_{SC}(T)$ , the significance of positive or negative values is inversed compared to  $\Delta E_{HB}(T)$  shown in fig. 5.6.



**Figure 5.8:** Difference between intra and inter-molecular side chain energies as a function of temperature. The top plot shows the difference  $\Delta E_{SC}(T) = E_{SC,intra}(T) - E_{SC,inter}(T)$ , while the bottom plot shows the derivative of  $\Delta E_{SC}(T)$ . Circles mark the transition temperatures  $T^*$  of the  $A\beta$  systems.

### 5.2.3 Hydrogen bond contact probability maps

Hydrogen bond contact probability maps (HB maps) are a powerful tool for visualizing the hydrogen bond networks formed within the different phases of the  $A\beta$  systems. The HB maps for intra-molecular HB contacts are displayed in fig. 5.9 and the HB maps for inter-molecular HB contacts are shown in fig. 5.10. It should be noted that the method of visualizing the HB maps of the  $A\beta$  systems differs from that of the HB maps of the PolyQ systems. While the HB maps of the PolyQ systems showed both intra- and intermolecular HB contacts in the same plot, the analysis of the  $A\beta$  systems utilizes separate HB maps for intra- or inter-molecular contacts. The different visualization was chosen to increase the HB map's readability and expressiveness. Since the  $A\beta$  systems are much larger than the PolyQ systems the number of beads on both axis of the HB maps is larger. Consequently, the cells in the HB maps are smaller, and the probability values are more densely packed, making it more difficult to distinguish between the different structures in the HB maps. Separation of intra- and inter contacts, reduces the number of beads per axis. Additionally, the  $A\beta$  systems exhibit more complex HB networks than the PolyQ systems and the differences between the maximal probabilities of intra- and inter-molecular HB contacts are larger. Therefore, structures appearing in the inter-quadrants of the HB map and that have lower contact probabilities compared to structures in the intra-quadrants are difficult to distinguish in the combined intra- and inter-molecular HB maps. Combined intra- and inter-molecular HB maps for the  $A\beta$  systems are provided in fig. C.2 in appendix C. For the separated representation of intra- and inter-molecular HB maps, the two corresponding intra- or inter-molecular quadrants of the combined HB maps were extracted and combined by averaging the values for each cell.

Another difference to the representation of HB maps in PolyQ systems is that the HB maps of the  $A\beta$  systems use a logarithmic probability scale. This scaling was chosen to better visualize the HB networks in the  $A\beta$  systems, which are more complex and have larger differences in

their probability values compared to the PolyQ systems. For comparison to the non-logarithmic HB maps, the combined intra- and inter-molecular HB maps of the  $A\beta$  systems are also provided using the logarithmic scale in fig. C.3 in appendix C.

As discussed in sections 3.4.3 and 4.2.2, HB maps provide a visual representation of the hydrogen bond networks formed within the systems and reveal the probability of HB contacts between pairs of residues in the system. The probability values are represented on a logarithmic scale, where red denotes high probability and blue indicates low probability. The color scale with the associated probabilities is provided on the right of the HB maps. Notably, the probability scales are consistent across all HB maps of the same interaction type (*intra* or *inter*). The maximum value on the color scale corresponds to the highest probability value across all displayed cells within the same figure (fig. 5.9 or fig. 5.10). The scales, however, differ between interaction types. To minimize numerical artifacts and focus on relevant data, all non-zero probability values below the threshold of  $10^{-3}$  are set to  $10^{-3}$ . Probability values of zero are represented in white, which applies exclusively to the white diagonal in the intra-molecular HB maps which represents the turn region where no HB contacts can form by definition of the PRIME20 model. Apart from the differences discussed above, the HB maps of the  $A\beta$  systems can be analyzed similarly to the PolyQ systems, i.e. secondary structure motifs in the peptide chains will produce the same kind of signatures in the different HB map representations of PolyQ and  $A\beta$ .

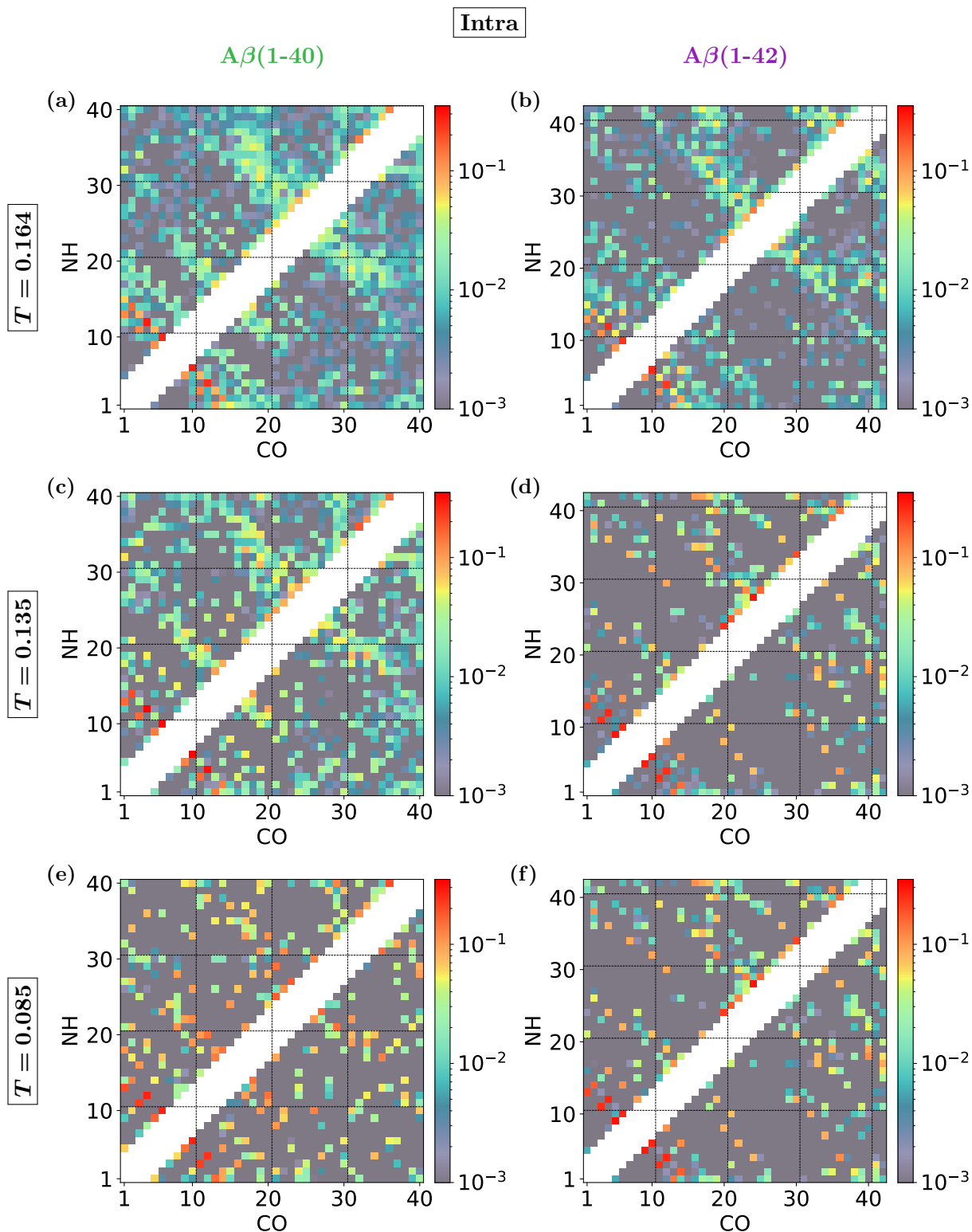
Figure 5.9 and fig. 5.10 display the HB maps of the  $A\beta$  systems at three selected temperatures. The temperatures were chosen to represent distinct phases of the systems. The first temperature ( $T = 0.164$ ) corresponds to the phase just below the first-order PT at  $T_{\text{high}}^*$ , while the second temperature ( $T = 0.135$ ) captures the phase below the second-order PT at  $T_{\text{low}}^*$ . The third temperature ( $T = 0.085$ ) was selected to illustrate the system's state at very low temperatures, below the low-temperature shoulder in the heat capacity  $C_V(T)$  of  $A\beta(1-40)$ , as well as below the structural features in  $\kappa^2$ ,  $N_{\text{HB,intra}}$  and  $N_{\text{HB,inter}}$  discussed in previous sections (see sections 5.1, 5.2.1 and 5.2.2). For temperatures above  $T_{\text{high}}^*$ , the  $A\beta$  systems remain unfolded with a very low probability of HB contact formation. Thus, they are of no further interest in the current analysis of the folded and aggregated states. For completeness, HB maps of the unfolded state at  $T = 0.195$  are included in fig. C.4 in appendix C.

For the remainder of this section, the notation  $(i, j)$  will designate the contact between the  $i$ -th and  $j$ -th residue in the sequence. The NH and the CO beads on the x- and y-axis are interchangeable in this context. When specifying a particular contact between  $N_i$  and  $C_j$ , the notation  $(N_i, C_j)$  will be used. Therefore, the notation  $(i, j)$  encompasses both HB contacts:  $(N_i, C_j)$  and  $(N_j, C_i)$ , thereby eliminating redundancy in the description of the HB interactions. Furthermore, the 3-letter abbreviations of the involved residues will occasionally be included in the  $(i, j)$  notation. This will be done to highlight specific interactions, such as the contacts at the turn of a  $\beta$ -hairpin.

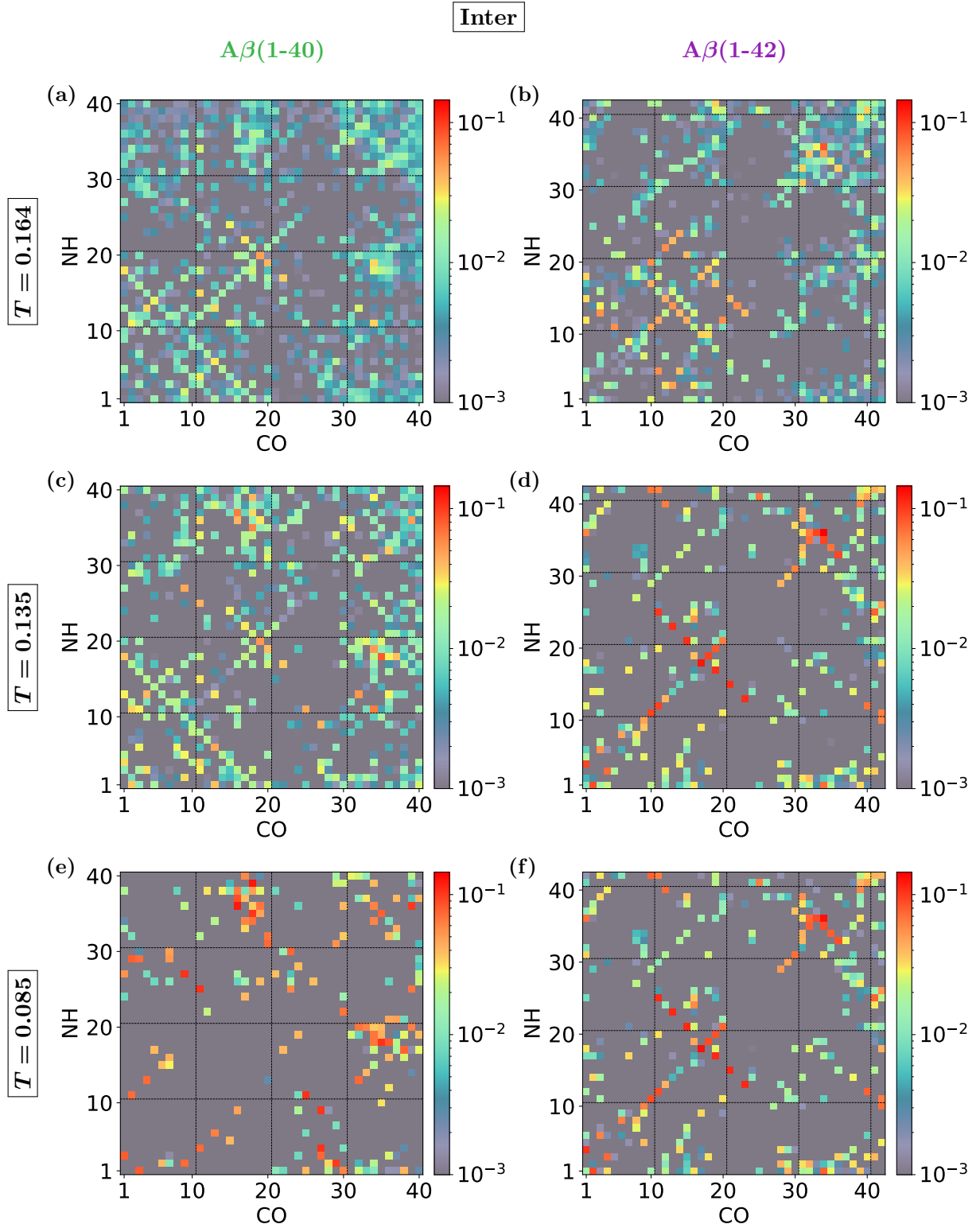
### Hydrogen bond networks at $T = 0.164$

Figure 5.9(a) and fig. 5.9(b) show the intra-molecular HB maps of  $A\beta(1-40)$  and  $A\beta(1-42)$  at  $T = 0.164$ , respectively. At first glance, both HB maps appear very similar. Both systems exhibit a high probability of HB contacts at the N-terminal region. Specifically, two  $\beta$ -hairpin signatures stick out. The hairpin structures will be identified by the residue contact that is closest to the turn. For the two  $\beta$ -hairpins, these *turn contacts* are (Arg5,Gly9) and (His6,Tyr10). Both have a high probability of occurring, with the latter being the more prominent in both systems. The two whole  $\beta$ -hairpin signatures involve the contacts  $[(5,9) - (3,11) - (1,13)]$  for the first  $\beta$ -hairpin signature and  $[(6,10) - (4,12) - (2,14)]$  for the second. Another common signature of both  $A\beta$  systems is the presence of an anti-parallel diagonal<sup>(6)</sup> region, centered around the turn contact

<sup>(6)</sup>The term 'anti-parallel diagonal' refers to the slope of the signature. As already discussed in the chapter on the PolyQ systems, the slope of the signature can be used to identify the orientation of the  $\beta$ -strands. In this case, the signature stretches from the top left to the bottom right, giving it a descending slope, which indicates



**Figure 5.9:** Intra-molecular hydrogen bond contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$ . The probability scale is logarithmic. The left column shows the  $A\beta(1-40)$  system, while the right column shows the  $A\beta(1-42)$  system. The first row displays the HBmaps at  $T = 0.164$ , the second row at  $T = 0.135$ , and the third row at  $T = 0.085$ .



**Figure 5.10:** Inter-molecular hydrogen bond contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$ . The probability scale is logarithmic. The left column shows the  $A\beta(1-40)$  system, while the right column shows the  $A\beta(1-42)$  system. The first row displays the HBmaps at  $T = 0.164$ , the second row at  $T = 0.135$ , and the third row at  $T = 0.085$ .

(Val24,Lys28) and identifiable by its bright yellow to green cells. Although, the probability of HB contacts in this region is lower compared to the N-terminal  $\beta$ -hairpin signatures, it still indicates an increased probability of anti-parallel alignment of the chain segments connected by (Val24,Lys28) or adjacent turn contacts. Secondary structure motifs that are consistent with this signature in the intra-molecular HB maps are different  $\beta$ -hairpins with different turn contacts.

Various configuration snapshots taken from the SAMC simulations of the  $A\beta$  systems are provided in fig. 5.11 for  $A\beta(1-40)$  and in fig. 5.12 for  $A\beta(1-42)$ . The configurations were selected for illustrative purposes and do not necessarily represent most probable system states. All configurations have energies equal to, or below, the average energy  $\langle U \rangle_T$  at temperatures below  $T_{\text{high}}^*$ . They therefore correspond to the folded and aggregated state at temperatures below  $T_{\text{high}}^*$ . In both figures, the chain's N-terminus is colored in blue and the C-terminus is colored in red. A consistent feature in all the presented configurations is the  $\beta$ -hairpin at the blue N-terminal end of the chain. Additionally, the region around the contact (Val24,Lys28) with yellow color often exhibits a turn connecting two  $\beta$ -strands in a  $\beta$ -hairpin, for example in figs. 5.11(b) and 5.11(h) for  $A\beta(1-40)$  and figs. 5.12(d) and 5.12(h) for  $A\beta(1-42)$ . These configurations are consistent with the anti-parallel diagonal signature observed in the HB maps that correspond to intra-molecular  $\beta$ -sheets and  $\beta$ -hairpins centered around the turn contact (Val24,Lys28) and adjacent turn contacts.

In  $A\beta(1-40)$ , the intra-molecular HB maps reveal a broader variety of possible intra-molecular HB contacts, compared to  $A\beta(1-42)$ . A prominent feature in  $A\beta(1-40)$  is the anti-parallel alignment around the turn contacts (His13,Leu17) and (His14,Val18). This signature indicates an increased probability of  $\beta$ -hairpin formation around these turn contacts and is much less pronounced in the HB map of  $A\beta(1-42)$ . Additionally, the  $A\beta(1-40)$  system shows a higher probability of HB contacts within the region  $([10 - 20], [20 - 40])$ , compared to the  $A\beta(1-42)$ 's HB map, where the probability values are much lower for these contacts. This region will be referred to as the *diversity area* in the remainder of this section. The *diversity area* in part shares contacts with the (Val24,Lys28)  $\beta$ -hairpin signature. However, it is difficult to assign a singular secondary structure to the *diversity area*, as the potential contacts are dispersed over a large area and do not form a distinct, singular signature. However, the overall antiparallel shape of many signatures in the *diversity area* suggests that intra-molecular  $\beta$ -sheets and  $\beta$ -hairpins are likely secondary structure candidates in this region.

In contrast, the intra-HB map of the  $A\beta(1-42)$  system exhibits a reduced range of possible intra-molecular HB contacts. While the signature of anti-parallel alignment around turn contact (Val24,Lys28) is still present in fig. 5.9(b), it is more confined and less spread out in the  $A\beta(1-42)$  system, indicating a more defined  $\beta$ -hairpin structure compared to the  $A\beta(1-40)$  system. The *diversity area* observed in the  $A\beta(1-40)$  system is notably less populated in the  $A\beta(1-42)$  system. The spread of low-probability intra-molecular HB contacts is significantly reduced, leaving fewer contacts but with higher probability that are compatible with  $\beta$ -hairpins. A reduced diversity is also noticeable in the regions  $([10 - 20], [20 - 30])$ , where there is a significantly lower probability of forming intra-molecular HB contacts. Instead, residue in this region are more likely to participate in inter-molecular HB contacts, as evident from the inter-molecular HB maps shown in figs. 5.10(a) and 5.10(b).

The inter-HB map of the  $A\beta(1-40)$  system at  $T = 0.164$  exhibits an overall lower probability of inter-molecular HB formation, compared to intra-molecular HBs. This was already evident in the HB contact analysis of the previous section (section 5.2.2). The most prominent feature is the (N<sub>20</sub>,C<sub>19</sub>) contact, which is part of an anti-parallel  $\beta$ -sheet signature. Furthermore, the HB map of  $A\beta(1-40)$  reveals several signatures of lower probability at the N-terminus, consistent with anti-parallel  $\beta$ -sheets (see fig. 5.11(a) for an example configuration). In contrast, the inter-HB map of the  $A\beta(1-42)$  system displays high-probability signatures of both parallel and

---

that the  $\beta$ -strands are oriented in an anti-parallel fashion.

anti-parallel alignment, particularly in the region  $([10 - 24], [10 - 24])$ , indicating a strong tendency for  $\beta$ -sheet formation between these beads (see figs. 5.12(c), 5.12(f) and 5.12(g) for example configurations). These signatures are significantly more pronounced than the corresponding ones in the  $A\beta(1-40)$  system. Additionally, the N-terminal region of the  $A\beta(1-42)$  system features a prominent anti-parallel  $\beta$ -sheet signature, spanning the contacts  $(N_1, C_{15}) \rightarrow (N_{15}, C_1)$ . This  $\beta$ -sheet signature has much higher probability than the corresponding N-terminal inter-molecular  $\beta$ -sheet signature in the  $A\beta(1-40)$  system. Example configurations, corresponding to this N-terminal  $\beta$ -sheet signature in  $A\beta(1-42)$  are shown in figs. 5.12(a) and 5.12(b). At the C-terminus, the intra-HB maps of both systems do not show any distinct signatures of specific secondary structures. However, the inter-HB map of the  $A\beta(1-42)$  system reveals a couple of high probability HB contacts at the C-terminus, most prominently the contact  $(N_{36}, C_{34})$ , which is near absent in the  $A\beta(1-40)$  system.

The analysis of intra- and inter-molecular HB maps at  $T = 0.164$  reveals significant differences between the two  $A\beta$  systems. The  $A\beta(1-40)$  system exhibits a wider array of possible intra-molecular HB contacts compared to the  $A\beta(1-42)$  system because HB signatures of notable probability are dispersed more broadly across the HB map for the  $A\beta(1-40)$  system, compared to  $A\beta(1-42)$ . This distribution of HB contact probabilities is an expression of the polymorphism of  $A\beta$  systems. Polymorphism is a well-known phenomenon in  $A\beta$  systems and has been described extensively for  $A\beta$  amyloid fibrils by numerous experimental and computational studies [3, 169–181]. In particular the exact position of the turn of the  $\beta$ -hairpins that constitute the core of the fibrils, as well as the specific interactions between residues in the intra- and intermolecular  $\beta$ -strands have been identified as key factors in the polymorphism of  $A\beta$  fibrils. Polymorphism in both  $A\beta$  variants has also been observed in monomer and dimer systems [173, 182–184], where folded and aggregated  $A\beta$  conformations vary in the structure of the N-Terminus and in the position of  $\beta$ -hairpin turns. The absence of a single, well defined folded and aggregated structure in the  $A\beta$  systems results in a spread out distribution of HB contacts in the HB maps shown here, since in the SAMC simulations, all configurations corresponding to a given temperature are considered simultaneously.

The intra-molecular HB maps reveal well defined  $\beta$ -hairpins at the N-terminus for both  $A\beta$  systems with two possible turn-positions: (Arg5,Gly9) and (His6,Tyr10). The center and C-terminal regions of the chains exhibit less probable contacts compared to the N-terminal  $\beta$ -hairpins. However, the shape of the signatures indicates  $\beta$ -hairpin formation in these regions as well, particularly around the turn contacts (Val24,Lys28) and (Asp23,Asn27). The spread of the probabilities in the signatures in the center and C-terminal half of the chains are indicative of the conformational polymorphism observed experimentally in both  $A\beta$  systems. The polymorphism is more pronounced in the  $A\beta(1-40)$  system, as evidenced by the broader range of possible HB contacts in the systems intra-HB map compared to those of the  $A\beta(1-42)$  system. The polymorphism is also visible in the inter-molecular HB maps, as evident by the spread of probable HB contacts. Furthermore, the  $A\beta(1-42)$  system shows well defined  $\beta$ -sheet signatures, particularly in the N-terminal half of the chains. Together with the overall higher probability for inter-molecular HB contacts, this reveals the more prominent role of aggregation in the PT at  $T_{\text{high}}^*$  for  $A\beta(1-42)$ , compared to  $A\beta(1-40)$ . On the other hand, the  $A\beta(1-40)$  system exhibits a broader range of possible HB contacts in the inter-HB map, emphasizing the more polymorphic nature of the  $A\beta(1-40)$  system compared to the  $A\beta(1-42)$  system, which was also visible in the intra-molecular HB maps. The HB map analysis further emphasizes the distinct folding and aggregation behavior of the two  $A\beta$  variants.

### Hydrogen bond networks at $T = 0.135$

Further changes to the HB networks in the  $A\beta$  systems are evident when the temperature is decreased to  $T = 0.135$ , below the second-order PT temperature  $T_{\text{low}}^*$ . Figures 5.9(c) and 5.9(d) show the intra-molecular HB maps of the  $A\beta(1-40)$  and  $A\beta(1-42)$  systems, respectively. At

$T = 0.135$ , the HB network of  $A\beta(1-40)$  closely resembles that at  $T = 0.164$ . The  $\beta$ -hairpin signatures at the N-terminus are still present, with the (His6,Tyr10) turn contact being the most prominent signature. The anti-parallel signatures around the turn contacts (His13,Leu17) and (His14,Val18) show an increased probability compared to  $T = 0.164$ . Within the *diversity area*, contact formation becomes more selective, with specific  $\beta$ -hairpin signatures becoming more pronounced. Additionally, the turn contacts ( $N_{24}, C_{20}$ ), ( $N_{25}, C_{21}$ ), ... ( $N_{29}, C_{25}$ ) show an increased probability, suggesting a greater likelihood of  $\alpha$ -helix formation. However, this probability remains significantly lower than that of the  $\beta$ -hairpin signatures at the N-terminus. In practice, configuration snapshots of  $A\beta(1-40)$  rarely display  $\alpha$ -helix formation. Therefore, these turn contacts likely do not form simultaneously to create whole  $\alpha$ -helices. Instead, they indicate single turns in the chain, stabilized by HB contacts. The probability of individual residues to be part of  $\alpha$ -helix structures will be further explored in section 5.2.5. Example configurations featuring  $\alpha$ -helix turns in the C-terminal half of the chain are shown in figs. 5.11(b) and 5.11(d).

The intra-molecular HB map of the  $A\beta(1-42)$  system undergoes much more pronounced changes, when the temperature decreases from  $T = 0.164$  to  $T = 0.135$ . The  $\beta$ -hairpin signatures at the N-terminus show an increase in probability, with the (Arg5,Gly9) turn contact and the (His6,Tyr10) turn contact becoming near equal in probability. Overall the range of possible HB contacts is significantly reduced compared to  $T = 0.164$ . However, the HB map reveals a variety of well-defined anti-parallel  $\beta$ -sheet signatures, such as [(2,26) – (4,24) – (6,22)], [(2,38) – (4,36)] and [(34,30) – (36,28) – (38,26)]. At  $T = 0.135$ ,  $A\beta(1-42)$  exhibits a much more specific and structured intra-molecular HB network compared to  $A\beta(1-40)$ . The beads [( $N_{24}, C_{20}$ ), ( $N_{25}, C_{21}$ ), ... ( $N_{29}, C_{25}$ )] and [( $N_{32}, C_{28}$ ), ( $N_{33}, C_{29}$ ), ( $N_{34}, C_{30}$ )] corresponding to residue contacts close to the turn display a high probability of forming HB contacts. Like in the  $A\beta(1-40)$  system, this again is consistent with signatures of  $\alpha$ -helix formation. However, as for the  $A\beta(1-40)$  system, configuration snapshots of the  $A\beta(1-42)$  system revealing  $\alpha$ -helix formation were rarely found. A rare example configuration snapshot of the  $A\beta(1-42)$  exhibiting an  $\alpha$ -helix loop at the C-terminal end of the chain can be seen in fig. 5.12(d).

The inter-molecular HB maps of the  $A\beta$  systems at  $T = 0.135$  are displayed in fig. 5.10(c) and fig. 5.10(d). The  $A\beta(1-40)$  system exhibits only minor changes compared to  $T = 0.164$ . Overall, the signatures become more defined, as low-probability signatures that were present at  $T = 0.164$  disappear. Notably, some short anti-parallel  $\beta$ -sheet signatures of higher probability appear, identified by their orange coloring. In contrast, the HB map of  $A\beta(1-42)$  changes much more significantly. The multitude of high-probability parallel and anti-parallel signatures in the region ([10 – 24], [10 – 24]) that were present at  $T = 0.164$  collapses into a singular anti-parallel  $\beta$ -sheet signature. This signature spans the residue contacts [(11,25) – (13,23) – (15,21)] and is the most probable signature in the inter-molecular HB map. Corresponding configurations are shown in figs. 5.12(c), 5.12(f) and 5.12(g). Additionally, the C-terminal region shows the high-probability contact ( $N_{36}, C_{34}$ ) that persists across the PT. Around it, more contacts with increased probability are revealed, forming an anti-parallel diagonal.

The changes in the HB networks of the  $A\beta$  systems at  $T = 0.135$  highlight notable differences in the folding and aggregation behavior of  $A\beta(1-40)$  and  $A\beta(1-42)$  at the second-order PT temperature. The  $A\beta(1-40)$  system exhibits only minor changes and again displays a broader array of possible intra-molecular HB contacts that is indicative of its greater degree of configurational polymorphism. The  $A\beta(1-42)$  system, on the other hand, shows significant change in its HB maps and features a more defined and specific intra-molecular HB network. A shared intra-molecular structural element in both  $A\beta$  systems is the persistent  $\beta$ -hairpin, which remains stable below the PT temperature  $T_{low}^*$ . The differences between the two  $A\beta$  systems are even stronger in the inter-molecular HB contacts.  $A\beta(1-42)$  exhibits a much higher probability of forming inter-molecular HB contacts and features a distinct  $\beta$ -sheet signature dominating the intra-molecular HB map. This indicates a more well-defined aggregated structure, than  $A\beta(1-$

40). In contrast,  $A\beta(1-40)$  shows a lower probability of inter-molecular HB contacts and a wider variety of potential HB interactions, reflecting the polymorphism of  $A\beta$  dimer systems, as was already present in the system at  $T = 0.164$ .

### Hydrogen bond networks at $T = 0.085$

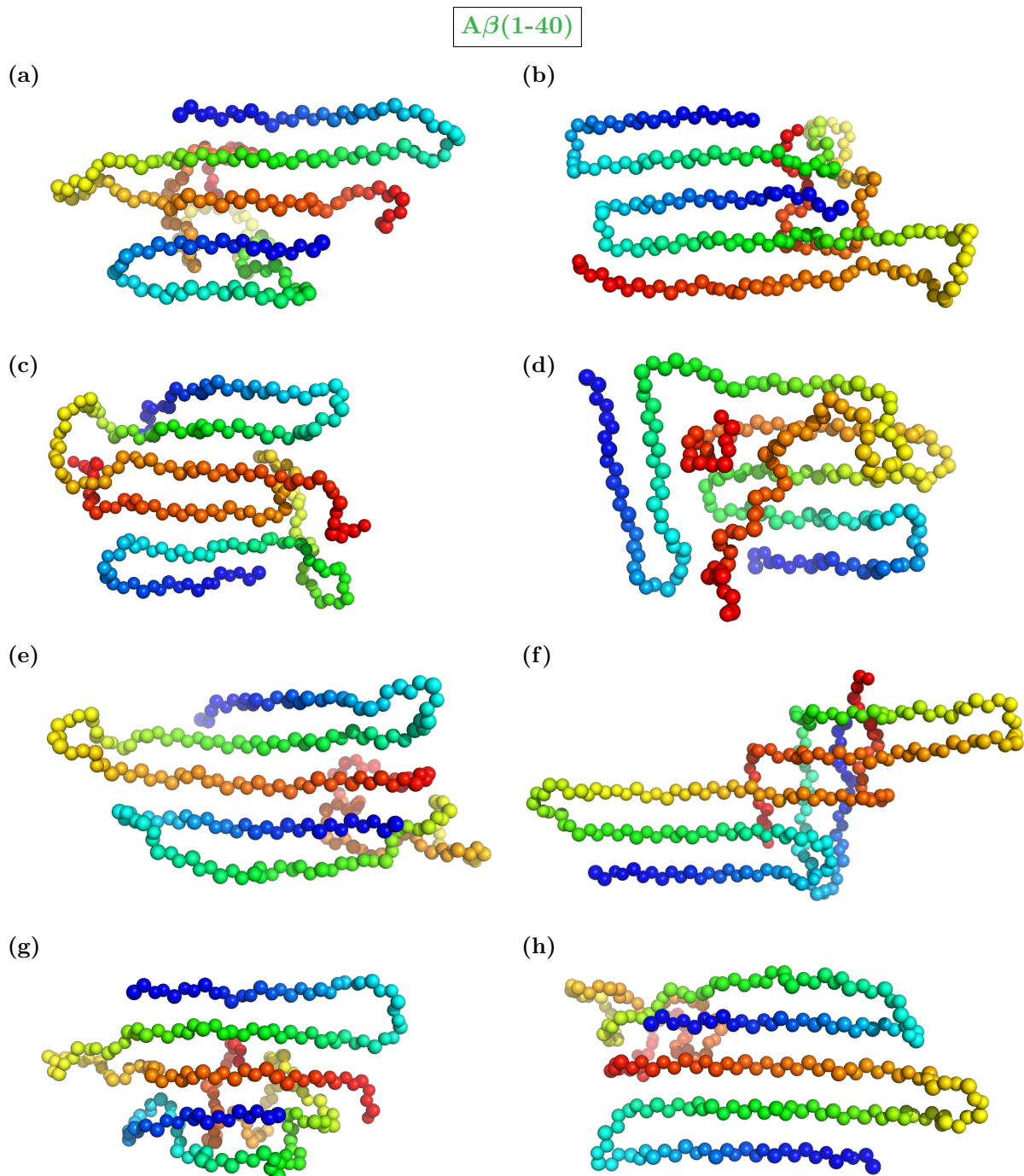
Finally, the bottom rows in figs. 5.9 and 5.10 show the HB map at  $T = 0.085$ . Even though this model temperature translates to  $T' = 150\text{K}$  in physical temperatures and is therefore far below physiological conditions, it still reveals interesting differences between the  $A\beta$  systems. In the intra-molecular HB map of  $A\beta(1-40)$  the diverse range of anti-parallel signatures collapse into specific, well defined  $\beta$ -hairpin signatures. Notably, the  $\beta$ -hairpins with turn contacts (His13,Val18), (His14,Val18) or (Asp23,Asn27) show a high probability of forming HB contacts. In contrast, the intra-molecular HB map of  $A\beta(1-42)$  does not change at all compared to  $T = 0.135$ . A similar behavior can be observed in the inter-molecular HB maps. In the  $A\beta(1-40)$  system the diversity of HB signatures is severely reduced, with only a few high-probability contacts remaining. For the most part, these contacts are in agreement with parallel and anti-parallel  $\beta$ -sheet formation. Notably, the large anti-parallel  $\beta$ -sheet signature at the N-terminus present at higher temperatures disappears, while only parallel  $\beta$ -sheet signatures remain. The  $A\beta(1-42)$  system, on the other hand, shows no change in the inter-molecular HB map compared to  $T = 0.135$ . The high-probability contacts at the C-terminus persist and the anti-parallel  $\beta$ -sheet signature in middle region remains stable.

In summary of the HB map analysis, the  $A\beta$  systems exhibit distinct folding and aggregation behaviors at different temperatures. Below  $T_{\text{high}}^*$ , both systems show a high probability of  $\beta$ -hairpin formation at the N-terminus. Signatures of further  $\beta$ -hairpins and anti-parallel  $\beta$ -sheets are present in the center and C-terminal region of chains in both  $A\beta$  systems. However, these signatures are spread out across multiple possible contacts belonging to multiple possible chain configurations. This is indicative of the polymorphism present in  $A\beta$  dimer systems [173, 183, 184].  $A\beta(1-40)$  system displays a much broader range of possible intra-molecular HB contacts compared to  $A\beta(1-42)$ , which indicates a higher degree of polymorphism in the  $A\beta(1-40)$  system.

The inter-molecular HB contacts in the  $A\beta(1-42)$  system are less spread out with higher individual probability and with clear signatures of  $\beta$ -sheet formation. In contrast, the  $A\beta(1-40)$  system shows a lower probability of inter-molecular HB contacts and a wider variety of potential HB interactions. This behavior again emphasizes the distinct aggregation behaviors of the two  $A\beta$  variants, with  $A\beta(1-42)$  favoring more clearly defined, specific interactions that have a much higher probability of occurring, compared to the more polymorphic nature of  $A\beta(1-40)$ .

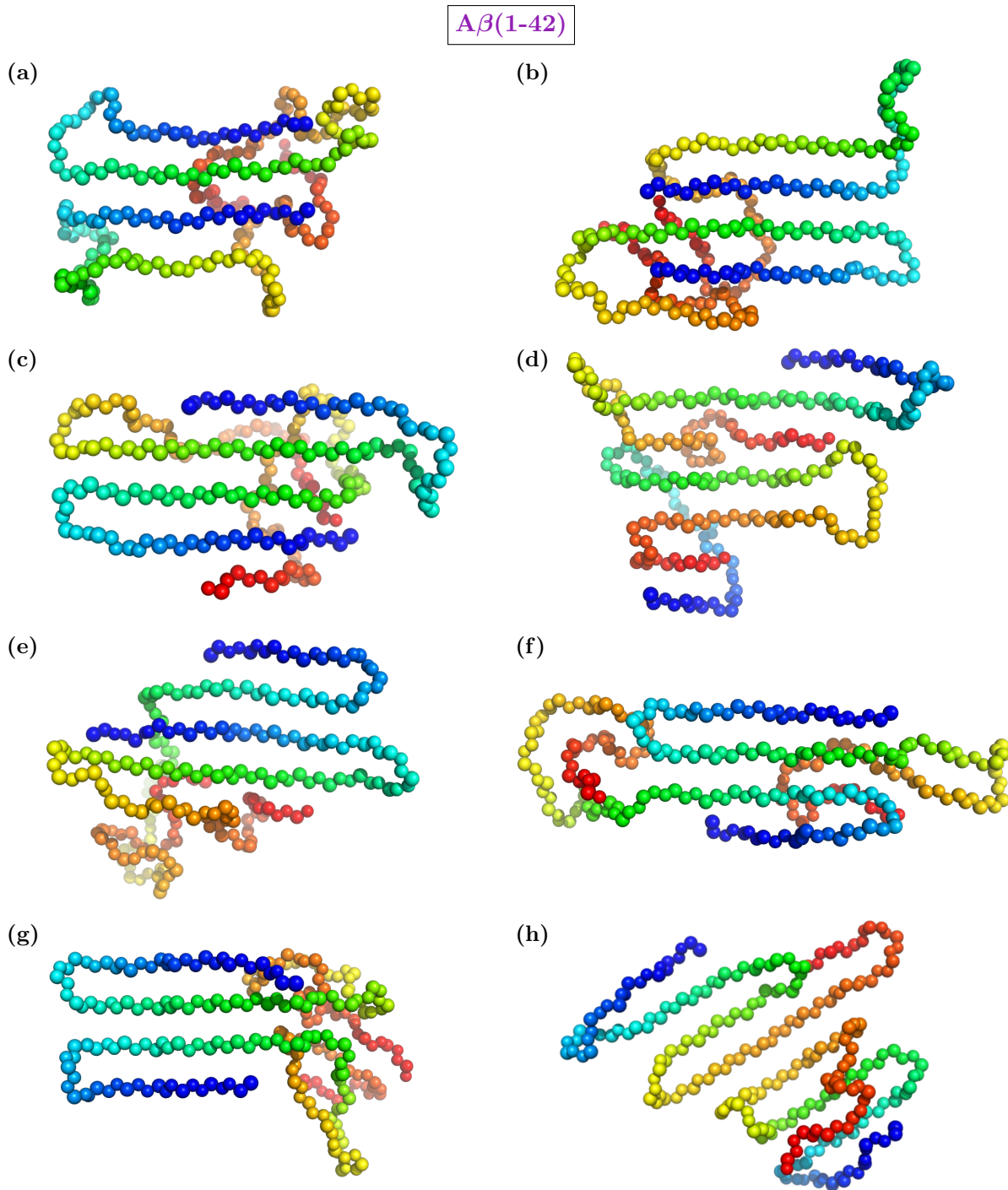
Below  $T_{\text{low}}^*$ , the HB network of the  $A\beta(1-40)$  system only changes in its inter-molecular HB contacts. The overall probability of inter-molecular HB contacts increases and specific HB contacts become more prominent. The  $A\beta(1-42)$  on the other hand, shows changes in both its intra- and inter-molecular HB network. The intra-molecular HB network becomes more defined, with specific  $\beta$ -hairpin and  $\beta$ -sheet signatures dominating the HB map. The signatures in the inter-molecular HB map show a well defined, high-probability  $\beta$ -hairpin signature involving residues in the middle region of the chain. The second-order PT in the  $A\beta(1-42)$  system is characterized by a fundamental change in the inter-molecular HB contacts, with a clear signature of  $\beta$ -sheet formation dominating the HB map.

It is not until the temperature is decreased to very low values, that the HB networks of the  $A\beta(1-40)$  system shows clearly defined  $\beta$ -hairpin signatures in its intra-HB maps and a reduction in the diversity of possible HB contacts. The inter-molecular HB map of the  $A\beta(1-40)$  system shows a similar reduction in diversity, with only a few high-probability contacts remaining. In contrast, the  $A\beta(1-42)$  system already shows well defined, aggregated secondary structures in the high-temperature aggregated phase (between  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ ). This includes room temperature  $T_{\text{R}}' = 300\text{K}$  which is equivalent to  $T_{\text{R}} = 0.170$  in model temperatures.



**Figure 5.11:** Configuration snapshots from simulations of the  $A\beta(1-40)$  system. The beads represent the backbone beads of the residues in PRIME20: NH,  $C_\alpha$  and CO. Side chain beads (R) are omitted for better visibility. The beads along the backbone of each of the two chains are colored in a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state.

The formation of  $A\beta$  dimers represents a crucial early step in the aggregation process of  $A\beta$  peptides, which eventually leads to the formation of larger oligomers and ultimately the amyloid fibrils. Both, oligomers and amyloid fibrils of  $A\beta$  are closely linked to the progression of AD. The HB networks observed in the simulated dimer systems show significant differences from those found in fibrillar structures. This paragraph compares the structures of  $A\beta$  fibrils and dimers reported in the literature with the folded and aggregated structures of the simulated dimer



**Figure 5.12:** Configuration snapshots from simulations of the  $A\beta(1-42)$  system. The beads represent the backbone beads of the residues in PRIME20: NH,  $C_\alpha$  and CO. Side chain beads (R) are omitted for better visibility. The beads along the backbone of each of the two chains are colored in a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state.

systems from this study. The HB maps used for this comparison are taken at temperatures that mimic physiological and experimental conditions. Specifically, HB maps at  $T = 0.164$  were chosen, representing the system in its high-temperature, folded and aggregated state between  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$  (figs. 5.9(a), 5.9(b), 5.10(a) and 5.10(b)).

The structure of  $A\beta$  fibrils has been studied extensively using nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM) [78, 171, 175, 185, 186]. A common result is

the identification of  $\beta$ -strands in the regions [12 – 24] and [30 – 40] that form inter-molecular HB contacts to create parallel inter-molecular  $\beta$ -sheets. In the center of the chain, between the two  $\beta$ -strands, a bend or turn is found, bringing the  $\beta$ -strands in anti-parallel alignment. These two  $\beta$ -strands of the same chain interact via SC interactions. The  $\beta$ -turn- $\beta$  motif is found in both  $A\beta(1-40)$  and  $A\beta(1-42)$  and constitutes the core of the amyloid fibril in a cross- $\beta$ -structure<sup>(7)</sup>.

The inter-molecular parallel  $\beta$ -sheets found in the amyloid-fibril of  $A\beta$  correlate to the parallel signatures found in the center region ([10 – 24], [10 – 24]) of the inter-molecular HB maps of  $A\beta(1-40)$ . Although, anti-parallel alignment appears more probable in the  $A\beta(1-40)$  dimers. The  $A\beta(1-42)$  dimers have a much higher probability of forming inter-molecular HB contacts in this region. Both parallel and anti-parallel  $\beta$ -sheet signatures are found in its inter-molecular HB maps. This implies, that the  $A\beta(1-42)$  system already shows the onset of fibril-like aggregation in its first-order PT. An example configuration adopting parallel alignment of inter-molecular  $\beta$ -strands in agreement with amyloid fibrils is displayed in fig. 5.12(g).

In the fibril, the loop which brings the peptide chains two  $\beta$ -strands into anti-parallel alignment is localized between residues 25-29 and stabilized by the salt bridge in the side chains (Asp23,Lys28) [78]. The two  $\beta$ -strands stabilize their anti-parallel alignment with intra-molecular SC contacts. In both simulated dimer systems, the chains also form  $\beta$ -strands in [12 – 24] and [30 – 40]. Furthermore, the intra-molecular HB maps revealed that these two  $\beta$ -strands form  $\beta$ -hairpins with intra-molecular HBs and a turn HB contact close to (Val24,Lys28). Thus, both  $A\beta$  dimer systems already exhibit  $\beta$ -strands in the same chain segments as in the full amyloid fibril. Furthermore, these  $\beta$ -strands in the dimer systems adopt anti-parallel alignment connected by the turn HB contact (Val24,Lys28), which is close with the salt bridge SC contact (Asp23,Lys28) found in the fibril. The main difference here is, that this hairpin structure is stabilized by HB contacts in the simulated dimer systems, whereas it is stabilized by SC contacts in the fibril.

The first 10 residues at the N-terminus of the  $A\beta$  peptides are found to be mainly unstructured in the fibril [78, 175, 185]. This is in stark contrast to the dimer systems simulated in this study, where the N-terminus shows a high probability of forming  $\beta$ -hairpins. This suggests that the formation of N-terminal  $\beta$ -hairpins plays a crucial role in the dimerization process, whereas it loses this importance in the fully formed fibril. Notably, studies by Bertini et al. [171] and Wälti et al. [186] found that the N-terminus of  $A\beta(1-40)$  and  $A\beta(1-42)$  peptides can indeed also adopt  $\beta$ -strand structures in the fibril. However, they are still not part of the cross- $\beta$  structure in the fibril-core.

NMR studies of  $A\beta$  peptides in aqueous solution are able to determine the structure of monomeric  $A\beta$  peptides [187–190]. Hou et al. [188] found, that both  $A\beta(1-40)$  and  $A\beta(1-42)$  adopt largely random, extended chain without a clearly defined secondary structure. However, the monomeric peptides show tendencies to form  $\beta$ -strands in the regions [17 – 21] and [31 – 36] as well as bend-like structures at regions [7 – 11] and [20 – 26]. The  $\beta$ -strand regions, together with the tendency to form turns around [20 – 26], align well with the  $\beta$ -hairpin signatures found in the intra-molecular HB map of the simulated dimer systems of this study. Furthermore, the turn at [7 – 11] is consistent with the  $\beta$ -hairpin signatures found at the N-terminus of the dimer systems.

Ball et al. [187] reported that  $A\beta(1-42)$  forms a prominent anti-parallel  $\beta$ -hairpin involving  $\beta$ -strands in the regions [16 – 21] and [29 – 36]. This is in agreement with the  $\beta$ -hairpin signatures observed in the intra-molecular HB map of the  $A\beta(1-42)$  system in this study. Additionally, Ball et al. found that  $A\beta(1-40)$  forms an alternative  $\beta$ -hairpin involving  $\beta$ -strands in the regions [9 – 13] and [16 – 21]. The analysis of the intra-molecular HB map for  $A\beta(1-40)$  in this study similarly shows a high probability of forming  $\beta$ -hairpins with turn contacts (His13,Leu17) or (His14,Leu17). In contrast, the probability of  $\beta$ -hairpin formation in this region is much lower in  $A\beta(1-42)$ . Therefore, the findings of this work are consistent with the structural differences

---

<sup>(7)</sup>A more thorough description of the cross- $\beta$ -structure can be found in section 2.1.1.

between the two A $\beta$  systems reported by Ball et al.

Next to experimental studies, simulations studies using MD and MC have also been used to study the structure of A $\beta$  peptides [169, 170, 183, 184, 191–194]. Anand et al. performed REMD simulations on monomer [191] and dimer [169] systems of A $\beta$ (1-39) at room temperature. They found that the monomeric peptide is unstructured and assumes random-coil conformations without any  $\beta$ -strand content. The dimer system, on the other hand, exhibits conformations with parallel N-terminals that were suggested to be stabilized by inter-molecular salt-bridges between SCs of Lys28 and Glu22 or Asp23. The systems of A $\beta$ (1-40) and A $\beta$ (1-42) in this study differ from the dimer structures reported by Anand et al. in that they show a preference for anti-parallel alignment and  $\beta$ -hairpin formation at the N-terminus. Furthermore, the salt-bridge between Lys28 and Asp23 does not appear in the HB maps as particularly important to the secondary structures of the A $\beta$  dimers in this study. However, the parallel alignment at the N-terminal, comparable to that found by Anand et al. is still a possible structure that is also found in conformations of the A $\beta$  dimers in this study (see for example fig. 5.12(g)). Furthermore, the length of the A $\beta$  peptide chains differs between the studies, which could explain the differences in the preferred alignment of the N-terminus. Another factor is the simulation technique used, with Anand et al. using REMD and this study using MC simulations, as the MC method allows for broader sampling of the conformation space.

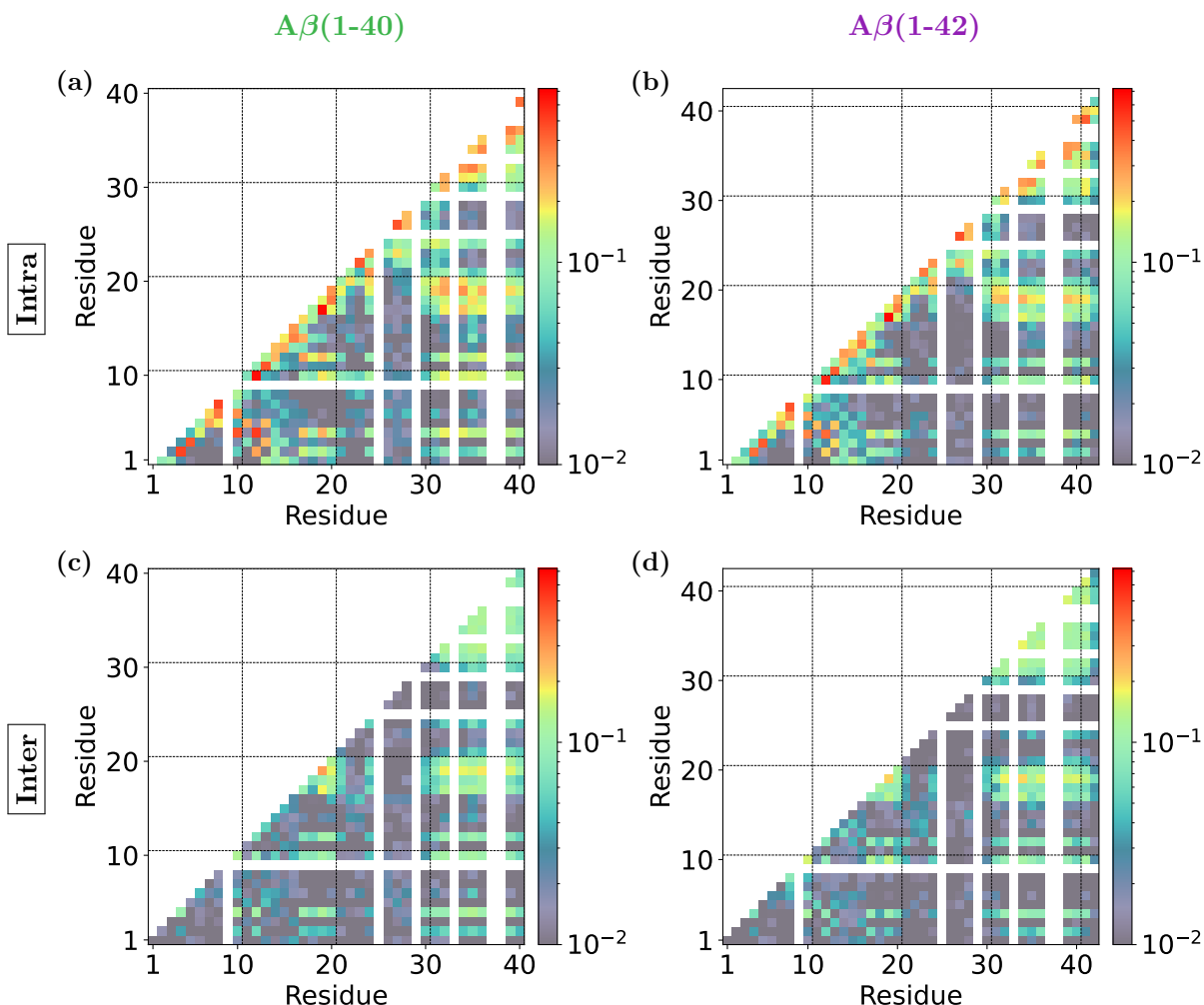
Côté et al. [184] performed REMD simulations on dimer systems of A $\beta$ (1-40) and A $\beta$ (1-42) at various temperatures. Their results show, that the N-terminal region of A $\beta$ (1-42) adopts very often a  $\beta$ -hairpin motif, centered at residues 7-9, while the N-terminal of A $\beta$ (1-40) populates mostly random coil configurations with a small fraction of intra- and inter-molecular  $\beta$ -sheets motifs. In contrast, the  $\beta$ -hairpin motif at the N-terminal end of the peptide chain is found with equal probability in both A $\beta$ (1-40) and A $\beta$ (1-42) in the dimer systems of this study. Côté et al. further found, that both A $\beta$  systems form  $\beta$ -strands in the regions [12 – 24] and [30 – 40]. Especially the region [12 – 24] was prone to form inter-molecular contacts, which is in agreement with results from the HB map analysis in this study.

#### 5.2.4 Side chain contact probability maps

To further understand the structural properties of the A $\beta$  systems, the probabilities of contacts between individual SC beads are analyzed. The side chain contact probability maps (SC maps) of the A $\beta$  systems at  $T = 0.164$  are shown in fig. 5.13. In contrast to the HB maps, the SC maps are symmetrical. This is due to each residue having only a single side chain, in contrast to two beads per residue (NH and CO) that can participate in HBs. This results in reciprocal contact probabilities and allows for diagonalization of the SC maps. The vertical and horizontal white lines in the SC map correspond to Gly residues, which lack the SC bead and therefore do not form SC contacts. Furthermore, since each SC bead can form multiple contacts with other SC beads simultaneously, the SC maps often display multiple high-probability contacts for individual beads.

**Inter**-molecular hydrogen bond contact probability maps for A $\beta$ (1-40) and A $\beta$ (1-42). The probability scale is logarithmic. The left column shows the A $\beta$ (1-40) system, while the right column shows the A $\beta$ (1-42) system. The first row displays the HB maps at  $T = 0.164$ , the second row at  $T = 0.135$ , and the third row at  $T = 0.085$ .

The intra-molecular SC contacts are shown in the top row of fig. 5.13. At  $T = 0.164$ , the SC map of the A $\beta$  systems are very similar. SC contacts between neighboring residues and next-neighboring residues show a high probability of forming contacts, which is to be expected due to their close proximity along the chain. Notably, at the N-terminal end of the chain, nearest neighbor residues show a lower probability than next-nearest neighbor residues. This is caused by the formation of  $\beta$ -hairpins at the N-terminus that was observed in the HB map analysis. The  $\beta$ -strand configuration causes neighboring SC beads to face in opposite directions of the chain, therefore neighboring SC beads do not interact in this configuration. However,

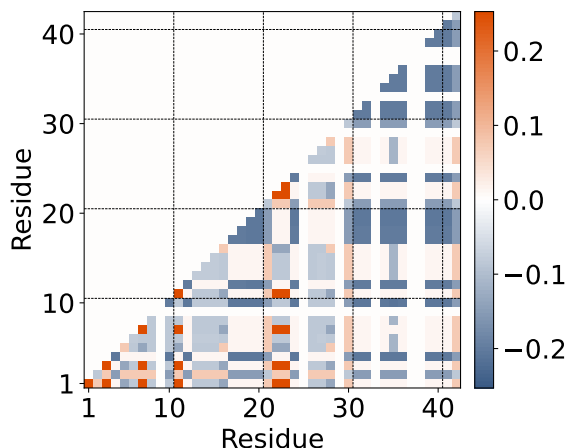


**Figure 5.13:** Intra- and inter-molecular side chain contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$  at  $T = 0.164$ . The color scale is logarithmic. The left column shows the  $A\beta(1-40)$  system, while the right column shows the  $A\beta(1-42)$  system. The top row displays the intra-molecular SCmaps, while the bottom row shows the inter-molecular SCmaps.

next-nearest neighbor SC beads again are on the same side of the  $\beta$ -strand and therefore show an increase probability of being in contact with each other. Additionally, there are two diagonal signatures of increased probability in both intra-molecular SC maps, involving residue contacts  $[(10, 6), (11, 5), (12, 4), (13, 3)]$  and the other  $[(8, 6), (10, 4), (12, 2)]$ . Equivalent to the HB maps, the diagonal shape indicates anti-parallel alignment of two  $\beta$ -strands at the N-terminus. In conjunction with the signatures close proximity to the turn region, this makes these signatures correspond to  $\beta$ -hairpin conformation.

In the C-terminal region of the chain, the SC maps show a high probability of SC contacts involving residues in the latter half of the chains. Specifically, intra-molecular SC contacts frequently form between residues in the range  $[30; 40(41)]$  <sup>(8)</sup> and residues  $[17; 20]$ , as well as  $[30, 40(41)]$ . Several factors contribute to the presence of these high-probability regions. First, the  $([30, 40(41)], [17; 20])$  contacts lie within the *diversity area* identified in the earlier HB maps analysis. This area is characterized by a high probability of HB formation and is therefore prone to feature SC interactions as well. Second, unlike the N-terminal region, which has a high likelihood to form specific  $\beta$ -hairpins, the C-terminal region lacks well-defined HB networks. This less structured nature of the C-terminus increases the probability and spread of SC contacts

<sup>(8)</sup>The number in round brackets refers to the position in the  $A\beta(1-42)$  system, while the number preceding the brackets corresponds to  $A\beta(1-40)$ .



**Figure 5.14:** Map of energy contributions of contacts between side chains of the residues of A $\beta$ (1-42) system when applying the PRIME20 force field. The color scale indicates the energy contribution of the contact, with blue colors representing attractive interactions and red colors representing repulsive interactions.

signatures due to the wider availability of potential contact partners. Furthermore, energy optimization of the chains influences the popularity of these contacts, as the energy contributions of the SC interactions depend on the specific amino acids involved, making certain contacts energetically more favorable. This becomes evident when plotting the potential energies of the individual SC contacts, as shown in fig. 5.14. Here, the SC energies defined by the PRIME20 model are mapped onto the different SC contacts in the A $\beta$ (1-42) protein. Blue colored contacts belong to attractive interactions and red ones to repulsive interactions, as shown on the color scale on the right. It immediately becomes apparent that the most probable contact partners for the SC beads of region [30; 40(41)], seen in figs. 5.13(a) and 5.13(b) are those SC beads with the most favorable energy contributions, which are colored in blue in fig. 5.14. The SC map patterns at the C-terminal end of the chain are therefore on the one hand a result of the less well established HB network which allows for more flexibility to form SC contacts, and on the other hand of the primary sequence of the A $\beta$  protein, which energetically favors SC contacts between beads in the C-terminal half of the polymer chain.

A similar argument applies to the inter-molecular SC contacts, shown in the bottom row of fig. 5.13. The probability pattern in the inter-molecular SC maps closely resemble those observed in the SC energy map in fig. 5.14, even more so than the intra-molecular SC maps. The key differences between the intra- and inter-molecular SC maps include a decreased probability of the diagonal signatures corresponding to  $\beta$ -sheets at the N-terminus. This reduction occurs because  $\beta$ -hairpins predominantly influence intra-molecular contact probabilities. Nevertheless, the anti-parallel signature faintly remains visible, which reflects the possibility of inter-molecular  $\beta$ -sheet formation in this chain segment. Examples for such configurations are shown in fig. 5.11(b) for A $\beta$ (1-40) and in fig. 5.12(a) for A $\beta$ (1-42). Additionally, the high-probability regions between beads  $(i, i - 1)$  and  $(i, i - 2)$  observed in the intra-molecular SC maps are absent in the inter-molecular SC maps, which is to be expected, since the SC beads belong to different chains and their proximity is not predefined by the primary sequence.

When decreasing the temperature below  $T_{\text{low}}^*$ , the SC contact probabilities of the A $\beta$  systems do not change significantly and are therefore not discussed here any further. SC maps of the A $\beta$  systems at these lower temperatures can be found in appendix C in figs. C.5 and C.6.

In conclusion, the analysis of the SC maps complements the findings of the previous analysis of the HB maps. The N-terminal end of the chain shows a clear preference for  $\beta$ -hairpin and inter-molecular  $\beta$ -sheet formation. The C-terminal end, on the other hand, is characterized by a greater diversity of possible SC contacts, which can be attributed to the lack of well-

defined HB networks. The energetically most favorable SC contacts in this region are the most commonly established contacts. The lack of a well defined HB network in the C-terminal region allows for greater flexibility in the configuration's selection of SC contacts. Notably, the difference in aggregation behavior between the two A $\beta$  systems that was observed in the HB maps are not obvious in the SC maps. Instead, the SC maps appear very similar for both systems. This suggests that the variation in the aggregation behavior of the A $\beta$  dimer systems is not caused by SC interactions, but rather by the HB interactions and their resulting HB networks. The formation of SC contacts is influenced by the primary sequence and the specific energy contributions of the SC interactions. While the system utilizes SC contacts for energy optimization, the energetically more significant HB contacts are the dominant factors that define the aggregation behavior of the A $\beta$  systems.

In the mature A $\beta$  fibrils, the SC interactions play a crucial role in stabilizing the cross- $\beta$  structure (see also section 2.1.1). The cross- $\beta$  spine consists of a pair of in-register, parallel  $\beta$ -sheets; each sheet is formed from  $\beta$ -strands that form HBs up and down the fibril axis. The  $\beta$ -sheet plane and the backbone HBs are oriented parallel to the fibril axis. The two  $\beta$ -sheets form an interface with a dry core, where the side chains of the residues in the  $\beta$ -strands intermesh in what is called a steric zipper [78, 170, 175, 177, 182, 195–197]. In the cross- $\beta$  structure of A $\beta$  fibrils, the two  $\beta$ -strands are formed in [12 – 24] and [30 – 40]. As established by the HB map and SC map analysis, these  $\beta$ -strands can also be found in the dimer systems simulated in this study. However, in the fibrils cross- $\beta$  structure, the two  $\beta$ -strands for a hairpin and interact via SCs, whereas in the dimer systems, the  $\beta$ -strands form a  $\beta$ -hairpin and interact via HBs. Whether the hairpin is stabilized by HBs or by SCs, the shape of the configuration in the fibril and the dimer system is similar in this chain region. Only the prioritization of the interaction type would be subject to change on the way from the dimer to the fibril.

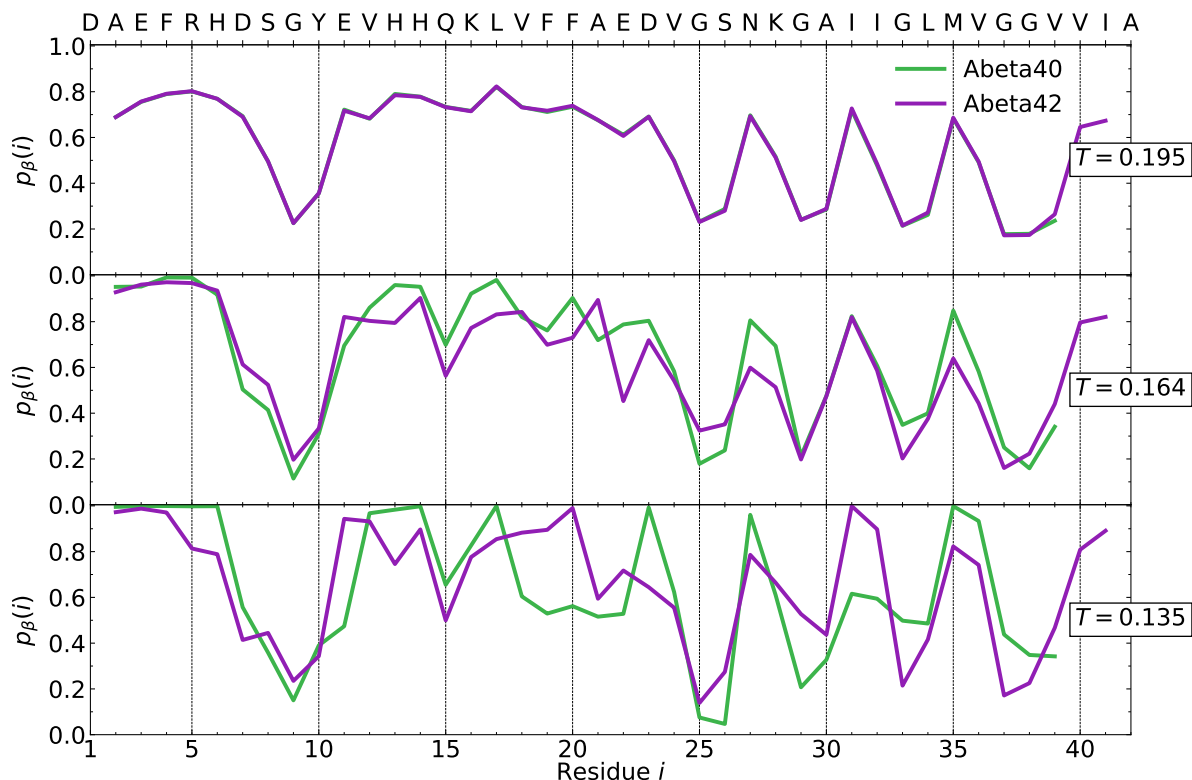
Another frequently reported feature of the A $\beta$  fibrils is the intra-molecular salt bridge between Asp23 and Lys28, which stabilizes the loop between the  $\beta$ -strands [78, 175]. This contact shows no special significance in the SC map analysis of the A $\beta$  dimer systems in this study. This is likely due to the fact that in the here simulated dimer systems, the SC interactions only play a minor role in the formation of the associated  $\beta$ -hairpin, compared to HB interactions. The apparent underestimation of the energy contribution of salt bridge in the associated hairpin might therefore indicate a limitation of the PRIME20 force field.

### 5.2.5 Ramachandran plot analysis

In addition to HB networks, the secondary structure of proteins can be characterized by the dihedral angles in the peptide backbone,  $\phi$  and  $\psi$ . The Ramachandran plot is a widely used tool for visualizing these angles. It enables the identification of angle combinations associated with specific secondary structures, such as  $\beta$ -strands and  $\alpha$ -helices. A more detailed discussion of Ramachandran plots can be found in sections 2.1 and 3.4. Here, the Ramachandran plot is used to identify the probability of a residue to be in a certain secondary structure state. The assignment of a secondary structure to the residue is performed for each residue individually, which allows for the determination of the probability distribution of specific secondary structure motifs along the protein chain. The residues at the two ends of the chain are excluded from the Ramachandran plot analysis, as they are missing one neighboring residue. Two neighboring residues are required for the calculation of  $\phi$  and  $\psi$ .

#### $\beta$ -sheet state

The probability distribution of  $\beta$ -strand associated dihedral angles  $p_{\beta}(i)$  over the whole protein chain is shown in fig. 5.15 for both A $\beta$  systems. Here,  $i$  is the residue index.  $p_{\beta}(i)$  is presented at 3 distinct temperatures, which were used in the preceding analyses as well, as they represent different phases of the A $\beta$  systems. The very low temperature state at  $T = 0.085$



**Figure 5.15:** Probability of the  $\beta$ -sheet state per residue along the chain for both A $\beta$  systems. The probabilities were calculated from the dihedral angles using the Ramachandran plot analysis. The probability is shown at three different temperatures,  $T = 0.195$ ,  $T = 0.164$ , and  $T = 0.085$ .

is not displayed in fig. 5.15 as this temperature is too low to be of physiological relevance and showed no significant change in  $p_\beta(i)$  (see section 5.1.1). Newly examined in this analysis is the temperature  $T = 0.195$ , which is above  $T_{\text{high}}^*$  and represents the phase where the A $\beta$  systems are in a disordered state. Notably, both A $\beta$  systems exhibit the same  $\beta$ -strand probability distribution at this temperature. Moreover, already in the disordered state, different areas of the A $\beta$  chains show different probabilities of  $\beta$ -strand formation, as is to be expected, given the different secondary structure propensities of the individual residues [198]. In the N-terminal half of the chain,  $p_\beta(i)$  generally rests at a value around 0.7 to 0.8, with minor deviations. It is important to note, that this does not directly correspond to the chain forming  $\beta$ -sheets in this region. The dihedral angles do not give any information about interconnectedness and proximity of residues. The high value of  $p_\beta(i)$  merely indicates that the dihedral angles of the residues in this region are likely to be oriented in a way that is characteristic for  $\beta$ -strands, i.e. the chain is likely to be elongated with neighboring side chains facing in opposite directions of the chain. A notable deviation from the high plateau of  $p_\beta(i)$  in the N-terminal half of the chain is the drop of  $p_\beta(i)$  at Gly9. Within the PRIME20 models geometry, this can be explained by the lack of a side chain in Gly residues, which allows for more flexibility in the backbone and therefore stronger deviations from the extended  $\beta$ -strand conformation. The C-terminal half of the chain shows a more complex behavior of  $p_\beta(i)$  in the disordered state, characterized by oscillations between high and low probabilities. In between the minima,  $p_\beta(i)$  reaches values around 0.7 which is similar to the values of  $p_\beta(i)$  in the plateau in the N-terminal half of the chain. The minima in  $p_\beta(i)$  are located at Gly25, Gly29, Gly33, Gly37 and Gly38. Again, the lack of a side chain in Gly and the resulting increased flexibility is likely the cause for the low probability of  $\beta$ -strand formation at these residues. Indeed, it is well-known that glycines in the protein sequence act as secondary structure breakers [74, 199–205]. The two additional residues in the A $\beta$ (1-42) system enable the calculation of the dihedral angles for Val40 and Ile41. These residues show a high

probability that their dihedral angles are in  $\beta$ -strand configuration.

When decreasing the temperature below  $T_{\text{high}}^*$ , to  $T = 0.164$ ,  $p_{\beta}(i)$  changes for both A $\beta$  systems in a similar way. Generally, previous tendencies that were present at the higher temperature above  $T_{\text{high}}^*$  are reinforced. The N-terminal half of the chain shows a high probability in its residues to have dihedral angles in  $\beta$ -strand configuration. The minimum at Gly9 remains persistent in both A $\beta$  systems. In the C-terminal half, the minima and maxima positions are conserved, but the  $p_{\beta}(i)$  values change and differ between the A $\beta$ (1-40) and A $\beta$ (1-42) system. The A $\beta$ (1-40) system shows more pronounced oscillations, while the A $\beta$ (1-42) system exhibits a decrease in  $p_{\beta}(i)$  at the maxima positions Asn27 and Met35. The two additional residues in the A $\beta$ (1-42) system increase in their probability of  $\beta$ -strand formation. Over all, the  $\beta$ -strand probability distribution is similar for both systems. The N-terminal half of the chain shows a consistently high probability of  $\beta$ -strand formation, with a significant drop at Gly9. This low-probability position corresponds to the turn region in the  $\beta$ -hairpins, identified in the previous analyses. The C-terminal half of the chain contains multiple residues that exhibit low  $\beta$ -strand probability. In between are residues of high  $\beta$ -strand probability. This complements the findings from the HB maps and SC maps, where the C-terminal half of the A $\beta$  chains was identified as a region of higher diversity and lower specificity in its secondary structure compared to the N-terminal half.

By cooling to  $T = 0.135$ , below the second order PT temperature  $T_{\text{low}}^*$ , the differences between the two A $\beta$  systems increase. For residues [18-22] the A $\beta$ (1-40) system shows a plateau with a value of  $p_{\beta}(i) \approx 0.6$ , while the A $\beta$ (1-42) system shows much higher values of  $p_{\beta}(i)$  in this chain region. This difference can be attributed to the A $\beta$ (1-42) system's increased probability of inter-molecular  $\beta$ -sheets. The HB maps in fig. 5.10(d) clearly identified inter-molecular  $\beta$ -sheets in the middle region of the chain at  $T = 0.135$  (see section 5.2.3). The high probability of  $\beta$ -strand configurations at the N-terminal persists for both A $\beta$  systems. In the C-terminal half,  $p_{\beta}(i)$  oscillates for both systems, although the variations are more pronounced in the A $\beta$ (1-42) system. Due to the inconsistent  $\beta$ -strand probabilities in the C-terminal half of the chain,  $p_{\beta}(i)$  does not give a clear picture of the  $\beta$ -strand formation in this region. Since the minima positions in  $p_{\beta}(i)$  of both systems are at, or at least remain close to, the positions of the Gly residues in the sequence, the lack of a clearly defined  $\beta$ -strand structure in the C-terminal half is likely due to the increased flexibility of the backbone in this region which allows for multiple turns and bends in the chain. However,  $p_{\beta}(i)$  is particularly low at residues Gly25 and Ser26 for both A $\beta$  systems, which fits well with the turn region of the  $\beta$ -hairpin signatures in the HB maps.

So far, the Ramachandran plot analysis of the A $\beta$  systems shows that the N-terminal half of the chain has a high probability of  $\beta$ -strand formation, with a drop at Gly9, which corresponds to the turn region of the  $\beta$ -hairpins. The C-terminal half of the chain shows more diverse behavior, with regions of high and low  $\beta$ -strand probabilities. The two additional residues in the A $\beta$ (1-42) system show a high probability of  $\beta$ -strand formation. The differences between the two A $\beta$  systems are most pronounced at low temperatures, where the A $\beta$ (1-42) system shows a higher probability of  $\beta$ -strand formation in the C-terminal half of the chain, and especially in the middle region [18 – 22]. This is likely due to the increased probability of inter-molecular  $\beta$ -sheet formation in the A $\beta$ (1-42) system, which is not present in the A $\beta$ (1-40) system. Furthermore, the minima in  $p_{\beta}(i)$  are located at the positions of the Gly residues, which do not have a side chain and therefore allow for greater flexibility in the backbone. Gly is known to break up and inhibit secondary structure formation [74, 199]. As there are 5 Gly residues dispersed throughout the C-terminal half of the chain, versus only 1 Gly residue in the N-terminal half, the primary sequence of the A $\beta$  systems likely contributes to the phenomenon that the N-terminal half of the chain has clearly defined  $\beta$ -hairpins in the folded and aggregated state, while the C-terminal half is more flexible and diverse in its secondary structure.

The distribution of  $\beta$ -strand probabilities in the A $\beta$  systems is in agreement with previous studies on A $\beta$  peptides. Bertini et al. [171] performed solid-state nuclear magnetic resonance

(ssNMR) experiments on fibrils of A $\beta$ (1-40) and predicted secondary structures from NMR chemical shifts using the TALOS+ program [206]. They found that residues [4 – 7] at the N-terminus are likely to form  $\beta$ -strands. They also found, that Ser8, Gly9 and Tyr10 show decreased  $\beta$  propensities, which could correspond to the turn region in the  $\beta$ -hairpin and is also observed in the Ramachandran plot analysis in this study. Bertini et al. also found high  $\beta$  probability in regions [11 – 23] and [31 – 40] at the C-terminus. The former high-probability region is in agreement with the Ramachandran plot analysis in this study, while the latter region does not show consistent high  $\beta$  probability. However, since Bertini et al. studied fibril systems, the region [31 – 40] is likely to be more ordered in their study compared to the dimer systems in this study.

Côté et al. [184] performed REMD simulations on dimer systems of A $\beta$ (1-40) and A $\beta$ (1-42) at various temperatures. They used the STRIDE program [207] to calculate secondary structure propensities. They found  $\beta$ -hairpin formation at the N-terminal end of the chain in both A $\beta$  systems to be the dominant morphology, while the C-terminal half of the chains was rather unstructured. This supports the findings of the Ramachandran plot and HB map analysis in this work. They also observed strong polymorphism in their structures, which is also present in the A $\beta$  systems of this work and has been discussed in the HB map analysis in section 5.2.3.

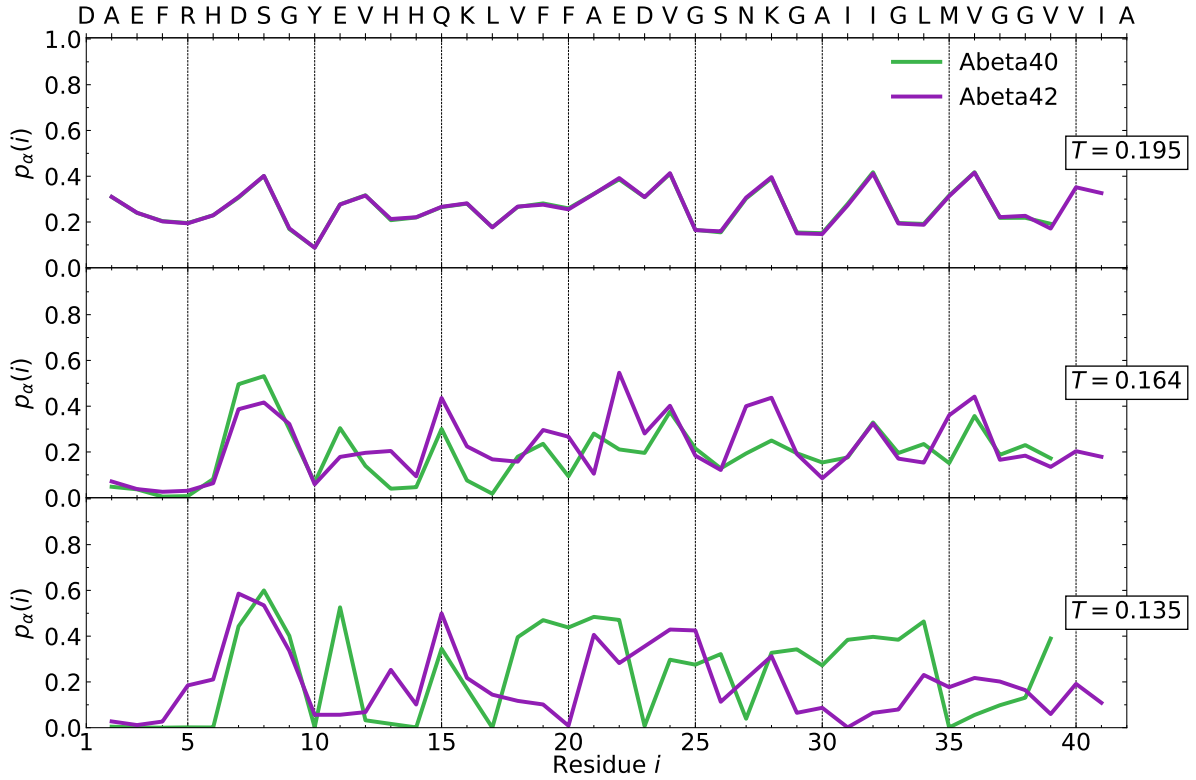
Urbanc et al. [181] performed DMD simulations on A $\beta$ (1-40) and A $\beta$ (1-42) oligomer systems. They also used the STRIDE program for the calculation of secondary structure propensities. In dimers, they found fluctuations in  $\beta$  propensities in the C-terminal half of the chain, with drops at the positions of the Gly residues, which is in agreement with the findings in this study. MC studies of A $\beta$ (1-42) monomers by Mitternacht et al. [193] also identified minima in  $\beta$  propensities at the positions of the Gly residues. However, the fluctuations in  $\beta$  propensities in the C-terminal half of the chain are more pronounced in the results of this study. This could be due to the SAMC sampling method using a flat-histogram approach, which allows for a more thorough exploration of the conformational space compared to the MC method used by Mitternacht et al. [193].

### $\alpha$ -helix state

Next to the  $\beta$ -strand probabilities, the Ramachandran plot analysis also provides the probability of  $\alpha$ -helix configuration of  $\phi$  and  $\psi$  for each residue. The probability distribution of  $\alpha$ -helix dihedral angles  $p_\alpha(i)$  over the whole chain is shown in fig. 5.16 for both A $\beta$  systems.

For temperatures above  $T_{\text{high}}^*$ , displayed in the top plot of fig. 5.16, the probability distribution of  $\alpha$ -helix formation is the same for both A $\beta$  systems and generally low. The minima in  $p_\alpha(i)$  are again at the positions of the Gly residues. The lack of SCs in the Gly residues and the accompanying increased flexibility in the backbone likely reduces the probability of the dihedral angle of Gly to be in  $\alpha$ -helix configuration, equivalent to the case with  $p_\beta(i)$ .

Differences between the A $\beta$  systems arise, when the temperature is lowered to  $T = 0.164$ , below  $T_{\text{high}}^*$ . A $\beta$ (1-42) shows more multiple peaks in  $p_\alpha(i)$  which are either not present or less pronounced in A $\beta$ (1-40). However,  $p_\alpha(i)$  needs to be consistently high across at least 4 residues in order to indicate whole  $\alpha$ -helix structures, because an  $\alpha$ -helix is defined by HB contacts between residues that are 4 residues apart in the sequence. This is not the case in the A $\beta$  systems, where the regions of higher values in  $p_\alpha(i)$  only span 1-2 residues. Therefore, the Ramachandran plot analysis does not show any clear  $\alpha$ -helix formation in the A $\beta$  systems. The low probability of  $\alpha$ -helix formation in the A $\beta$  systems is consistent with the findings from the HB maps, where no clear  $\alpha$ -helix structures were identified at  $T = 0.164$ . Another noteworthy feature in  $p_\alpha(i)$  of both A $\beta$  systems is the very low  $\alpha$ -helix probability at the N-terminal end of the chain and the increased probability around the Ser8 residue. This is consistent with the simultaneous drop in  $p_\beta(i)$  around the Gly9 residue, where the turn region of the  $\beta$ -hairpins is located. The increase in  $p_\alpha(i)$  indicates that the dihedral angles in the turn region of the  $\beta$ -hairpins have a tendency to adopt  $\alpha$ -helix configurations to achieve the necessary turn in the chain.



**Figure 5.16:** Probability of the  $\alpha$ -helix state per residue along the chain for both A $\beta$  systems. The probabilities were calculated from the dihedral angles using the Ramachandran plot analysis. The probability is shown at three different temperatures,  $T = 0.195$ ,  $T = 0.164$ , and  $T = 0.085$ .

The  $\alpha$ -helix probability distributions of both A $\beta$  systems undergo further change, when decreasing the temperature to  $T = 0.135$ , below  $T_{\text{low}}^*$ . In the A $\beta$ (1-40) system, two plateaus of  $p_\alpha(i) \approx 0.4$  to  $0.5$  are visible at residue range  $[18, 22]$  and at  $[28, 34]$ . With these regions of higher  $p_\alpha(i)$  spanning more than 4 residues, these regions in the chain correspond to the potential  $\alpha$ -helix signatures found in HB map analysis at  $T = 0.135$  (see section 5.2.3). The plateau in range  $[18, 22]$  matches the region of low probability in  $p_\beta(i)$  of the A $\beta$ (1-40) system. However,  $p_\alpha(i)$  remains below  $0.5$  which indicates that even the regions of high probability of  $\alpha$ -helix formation are still low compared to the  $\beta$ -strand formation probabilities. Furthermore, configurations featuring  $\alpha$ -helices are rarely found in simulation snapshots. An example of a A $\beta$ (1-40) configuration with a potential  $\alpha$ -helix around the Ala30 residue is shown in fig. 5.11(d).

The A $\beta$ (1-42) system shows a higher-value plateau in  $p_\alpha(i)$  in the residue range  $[21, 25]$  with  $p_\alpha(i) \approx 0.4$ . As in the A $\beta$ (1-40) system, this indicates potential  $\alpha$ -helix formation in this residue range. However, the probability again is low, compared to the  $\beta$ -strand formation and  $\alpha$ -helix signatures are uncommon in the simulation snapshots. Indeed, the configurations featuring  $\alpha$ -helices shown in figs. 5.12(d) and 5.12(h) both show helical structure motifs around the Met35 residue, instead of the potential  $\alpha$ -helix region around residues  $[21, 25]$ . However, the selection of these snapshots is highly influenced by random selection and by visual expressiveness. The snapshots therefore do not provide a reliable representation of the most probable structural features of the A $\beta$  systems and merely serve illustrative purposes.

In conclusion, the Ramachandran plot analysis of the A $\beta$  systems does not show any clear  $\alpha$ -helix formation. The N-terminal region of the chain shows a low probability of  $\alpha$ -helix formation, with a slight increase around the turn region of the  $\beta$ -hairpins. At very low temperatures, potential  $\alpha$ -helix signatures are visible in the Ramachandran plot analysis, but these are not consistently high and are not reflected in the simulation snapshots. The regions of these  $\alpha$ -helix signatures differ between the two A $\beta$  systems and match signatures of  $\alpha$ -helices in the HB maps

analysis.

Different studies have described helix formation in monomeric  $A\beta$  systems [208–210]. However, the Ramachandran plot analysis in this study does not show any clear  $\alpha$ -helix formation in the dimer systems. A reason for this discrepancy could be the solvent modeling in the PRIME20 model. The PRIME20 model uses implicit solvent and does not model solvent molecules explicitly. Specific solvent-peptide interactions that potentially influence the peptides propensity for  $\alpha$ -helix formation are averaged out in the PRIME20 model.

### 5.3 Conclusions

Using SAMC and the PRIME20 protein model, the thermodynamic and structural properties of the  $A\beta(1-40)$  and  $A\beta(1-42)$  dimer systems were investigated. Analysis of the microcanonical and canonical heat capacities revealed distinct phase transitions occurring at  $T_{\text{low}}^*$  and  $T_{\text{high}}^*$  for each system. At  $T_{\text{high}}^*$ , the systems undergo a first-order PT, while at  $T_{\text{low}}^*$ , a second-order PT is observed. The specific PT temperatures are provided in table 5.1. Furthermore, analysis of the intra- and inter-molecular HBs as a function of temperature revealed, that the high-temperature transition corresponds to a simultaneous collapse and aggregation of the  $A\beta$  chains. The low-temperature transition involves a restructuring of the HBs within the system.

In order to relate the results of the simulations to physiological and experimental conditions, the conversion factor between reduced model units and physical units was determined (see eq. (5.1)). It was found that room temperature lies between the two PT temperatures of the  $A\beta$  systems. Thus, both  $A\beta$  systems are in their aggregated states at room temperature. This is consistent with the known aggregation behavior of  $A\beta$  peptides, which form fibrils at physiological temperatures. However, thermodynamic analysis of the systems using the microcanonical entropy  $S(E)$ , which is the primary result of the SAMC simulations, revealed significant differences in the folding and aggregation behavior of  $A\beta(1-40)$  and  $A\beta(1-42)$ . These distinctions were also apparent in the structural analysis of the  $A\beta$  systems. Comparing intra- and inter-molecular HB and SC contacts showed that aggregation plays a significantly larger role in the PTs of the  $A\beta(1-42)$  system compared to  $A\beta(1-40)$ .

Contact probabilities between individual beads were analyzed using HB maps and SC maps, providing detailed insights into the ensemble of structures adopted by the  $A\beta$  systems at different temperatures. The results complement previous *in vitro* and *in silico* studies on fibril, oligomer and monomer structures of  $A\beta$  peptides, while offering new insights into the aggregation behavior of  $A\beta$  dimers. The HB map analysis of the aggregated state revealed a high probability of  $\beta$ -hairpin formation in the N-terminal region of both  $A\beta$  systems, with the Gly9 residue in the center of the turn. Furthermore, the systems showed a tendency to form  $\beta$ -hairpins around turn contact (Val24,Lys28). However, this region exhibited a less well-defined HB network compared to the N-terminal hairpins. The observation that  $A\beta$  exhibits multiple possible folded and aggregated configurations is an expression of the polymorphic nature of  $A\beta$  peptides, a well-known phenomenon observed in fibril as well as monomer structures across various experimental and computational studies [169–181]. The HB maps revealed greater diversity in the HB networks of  $A\beta(1-40)$ , than in  $A\beta(1-42)$ . Notably, at room temperature,  $A\beta(1-42)$  already displays a well-defined aggregated structure involving the middle region of the chain, while  $A\beta(1-40)$  did not. This finding is consistent with the known phenomenon, that  $A\beta(1-42)$  forms fibrils more rapidly than  $A\beta(1-40)$  [24]. The more well-defined aggregated structure of  $A\beta(1-42)$  would benefit faster aggregation kinetics.

Although both  $A\beta$  systems adopt aggregated structures reminiscent of the steric-zipper motif found in  $A\beta$  fibrils, the aggregated dimer was stabilized by HBs rather than SCs. However, this potentially represents an intermediate stage towards fibril formation. Furthermore, parallel alignment of the chains in the dimer systems is also observed, which may represent an initial step in the formation of parallel  $\beta$ -sheets characteristic of the cross- $\beta$  fibril structure. However, this

parallel alignment did not show distinct preference over anti-parallel alignment in the aggregated structures at room temperature.

The analysis of the Ramachandran plots reinforced the finding that the N-terminal region of the chains in the dimer systems has a high probability of  $\beta$ -hairpin formation, with a pronounced drop in probability around the Gly9 residue at the turn. In contrast, the C-terminal half of the chains exhibited more variability in the  $\beta$ -strand probabilities, with alternating regions of high and low probabilities. Notably, the Gly residues in the C-terminal region correspond to minima in the  $\beta$ -strand probabilities. Given that Gly is a known secondary structure breaker, the distribution of Gly residues over the amino acid sequence of A $\beta$  likely contributes to the absence of a well-defined secondary structure in the C-terminal half of the A $\beta$  chain. Furthermore, the Ramachandran plot analysis revealed a low probability of  $\alpha$ -helix formation across the A $\beta$  systems.

The findings in this chapter demonstrate that while the A $\beta$ (1-40) and A $\beta$ (1-42) dimer systems share similarities in their secondary structures, they exhibit significant differences in their quaternary structure, particularly in their aggregation behavior. Specific regions in the A $\beta$ (1-42) system that promote structure formation along the pathway to mature amyloid fibrils were identified. These results align with previous experimental and computational studies on A $\beta$  peptides, while offering valuable new insights into the aggregation behavior of the A $\beta$  dimer systems. This study enhances our understanding of the phase behavior of A $\beta$  peptides over a large temperature range, an essential step towards developing therapeutic strategies for Alzheimer's disease.

## Chapter 6

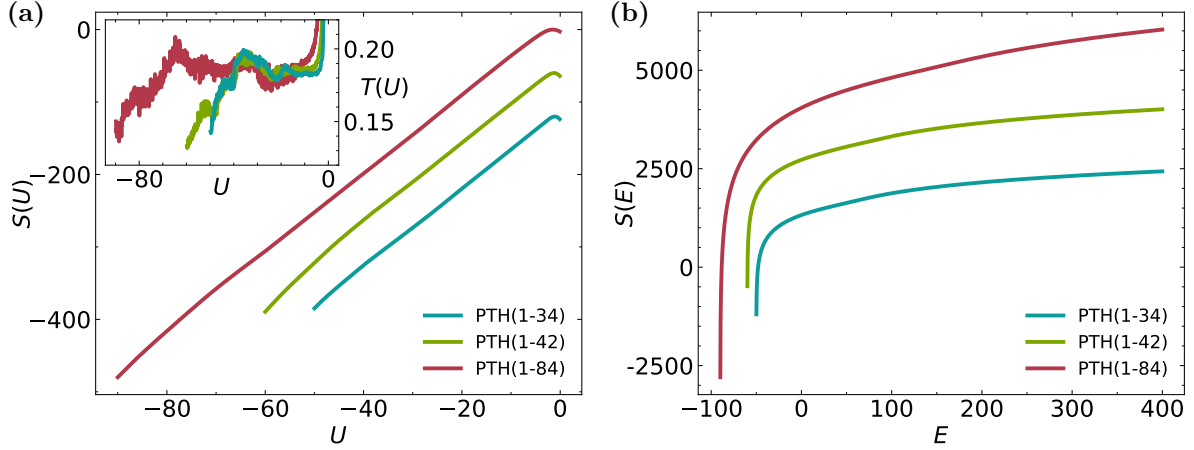
# Dimerization of Human Parathyroid Hormone

The two previous chapters examined peptides crucial to the development of neurodegenerative diseases, specifically Huntington’s disease and Alzheimer’s disease. This chapter will focus on a peptide that is not directly linked to a particular pathology, but is of high biological relevance. Parathyroid hormone (PTH) is a peptide hormone that plays a central role in calcium homeostasis. It is secreted by the parathyroid glands and acts on the parathyroid hormone receptor (PTHr) in the kidneys, bones, and intestines. The primary function of PTH is to increase the concentration of calcium in the blood. Although PTH is a peptide of 84 amino acids, only the 34 amino acids at the N-terminal end are required for its biological activity [211–213]. Similar to  $A\beta$  and PolyQ, PTH can form amyloid fibrils under certain conditions. However, these PTH amyloid fibrils are not associated with disease but rather represent a physiological process where the fibril structure serves as a storage form of the hormone.

In this chapter, the structural and thermodynamic properties of PTH dimer systems will be investigated. Three different PTH fragments are considered: the full length 84 amino acid peptide PTH(1-84), the N-terminal 42 amino acid peptide PTH(1-42) for a comparable length to  $A\beta$ , and the N-terminal 34 amino acid peptide PTH(1-34) which is the biologically active fragment. The analysis focuses on the temperature dependent phase behavior, the HB interactions between residues and the resulting secondary structure of PTH.

### 6.1 Thermodynamic analysis

The entropy  $S(U)$ , approximated using the SAMC simulations, is presented in fig. 6.1(a) for all three PTH dimer systems. The entropy curves of each system exhibit a similar shape with an overall ascending slope. For larger systems with longer chains,  $\ln g(U)$  converges at lower energies compared to shorter chains. This is due to the increased number of amino acids in longer chains, which enables the formation of more energetically favorable contacts between beads. Notably,  $S(U)$  reaches a maximum below the highest energy value,  $U_{\max} = 0$ , which is an effect attributed to the finite size of the system and the specific geometric constraints of the PRIME20 model. For a more detailed discussion of this phenomenon, the reader is referred to sections 4.1 and 5.1, as the same explanation applies to the PTH systems. Although the  $S(U)$  graphs appear smooth, they display noise due to the two energy scales in the PRIME20 model. This noise becomes more evident when analyzing the temperature  $T(U)$ , calculated as the inverse of the derivative of  $S(U)$  according to eq. (3.37). The temperature  $T(U)$  is shown in the inset graph of fig. 6.1(a). The noise in  $T(U)$  makes further thermodynamic analysis of the system difficult. To address this issue and to furthermore achieve a more accurate representation of the microcanonical ensemble, it is necessary to account for the contributions of the kinetic energy. The conversion from the configurational to the full microcanonical ensemble ( $S(U) \rightarrow S(E)$ ) is



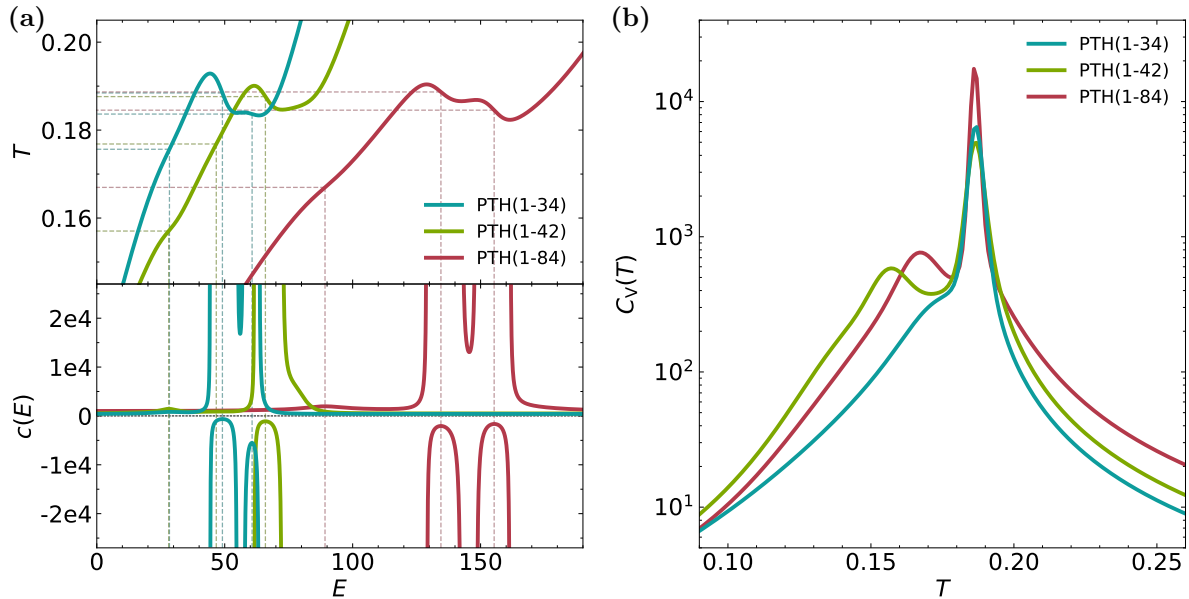
**Figure 6.1:** Configurational and total-energy microcanonical entropy of PTH(1-34), PTH(1-42) and PTH(1-84). (a) Entropy  $S(U)$  as a function of potential energy  $U$ . Inset: Microcanonical temperature  $T(U)$ . (b) Entropy  $S(E)$  as a function of total energy  $E$ .

discussed in section 3.4 and has been applied in the analysis of the PolyQ and A $\beta$  systems. The resulting  $S(E)$  graphs are displayed in fig. 6.1(b).

In order to analyze the shape of  $S(E)$ , the temperature  $T(E)$  and the microcanonical heat capacity  $c(E)$  are examined. These quantities are derived from the derivatives of  $S(U)$  as described in eqs. (3.37) and (3.38) and are displayed in fig. 6.2(a). As discussed in the previous two chapters, maxima in the microcanonical heat capacity indicate PTs at specific energy levels  $E_{\max}$ . A negative value of  $c(E_{\max})$  corresponds to a negative slope of  $T(E)$  at  $E_{\max}$ , which is indicative of a first-order PT. Conversely, a positive value of  $c(E_{\max})$  corresponds to a positive slope of  $T(E)$  at  $E_{\max}$  which indicates a second-order PT. The resulting PT temperatures are summarized in table 6.1. PTH(1-34) and PTH(1-84) exhibit two first-order maxima in  $c(E)$ , whereas PTH(1-42) shows only one. However, the two maxima for PTH(1-34) and PTH(1-84) are part of the same large scale fluctuation in  $T(E)$ . As such, they are considered as belong to the same PT. The temperature of this PT can be determined via a Gibbs-construction. The PT temperatures obtained through this method are also listed in table 6.1. The first-order PT temperatures for all three systems are approximately  $T_{\text{high}}^* \approx 0.187$ . Additionally, all 3 systems exhibit peaks corresponding to second-order PT in their  $c(E)$  graphs, with the respective transition temperatures  $T_{\text{low}}^*$  occurring below the first-order PT temperatures.

**Table 6.1:** Phase transition temperatures  $T^*$  of the three PTH systems, obtained via microcanonical and canonical analysis.

System	$T_{\text{micro}}^*$			$T_{\text{can}}^*$
	1 <sup>st</sup> order	2 <sup>nd</sup> order	Gibbs-loop	
PTH(1-34)	—	0.1756	—	—
	0.1837	—	—	—
	0.1885	—	0.1867	0.187
PTH(1-42)	—	0.1570	—	0.157
	—	0.1768	—	—
	0.1876	—	0.1867	0.187
PTH(1-84)	—	0.1670	—	0.167
	0.1845	—	—	—
	0.1887	—	0.1864	0.186



**Figure 6.2:** (a) Microcanonical temperature  $T(E)$  and heat capacity  $c(E)$ . The positions of local maxima in  $c(E)$  are marked by dashed lines and transferred to  $T(E)$  to find the corresponding PT temperatures. (b) Canonical heat capacity  $C_V(T)$ .

To complement the microcanonical analysis, the canonical heat capacity  $C_V(T)$  is examined, as shown in fig. 6.2(b). The sharp peaks that are present for all 3 PTH systems around  $T \approx 0.187$  indicate first-order PTs. These are consistent with the transition temperatures  $T_{\text{high}}^*$  identified in the microcanonical ensemble via the Gibbs-construction. Smaller, less sharp peaks are visible at lower temperatures for PTH(1-42) and PTH(1-84) and indicate second-order PTs. While the second-order transitions are clearly visible for the two larger systems, the one that was identified in the microcanonical analysis of PTH(1-34) appears only as a slight shoulder in the canonical heat capacity, and is lacking a distinct maximum. All PT temperatures determined in the canonical analysis are summarized in the last column of table 6.1.

The thermodynamic analysis reveals some key similarities as well as major differences between the three PTH systems. Notably, the temperature of the first-order PT is independent of the chain length, with  $T_{\text{high}}^* \approx 0.187$  for all three systems. This is somewhat surprising, as the folding transition of a polymer chain usually increases with the chain length. The PolyQ systems studied in chapter 4 serve as an excellent example of this behavior. However, in contrast to PolyQ, which is a homopolymer composed solely of glutamine residues, PTH is a heteropolymer. The sequence of PTH(1-84) consists of 18 different amino acid residues. This heterogeneity likely leads to distinct folding behavior along the chain which influences the overall transition temperature.

Furthermore, as will be detailed in the structural analysis sections that follow, the PT at  $T_{\text{high}}^*$  is not merely a folding transition but also involves aggregation. This is analogous to the behavior observed for the PolyQ and A $\beta$  systems (see chapters 4 and 5). In the PRIME20 model, both aggregation and folding occur on the same energy scale, driven by the energy optimization through HBs, with  $E_{\text{HB}} = -1$ . Consequently, both transition take place at the same temperature. In contrast to the folding transition, which depends on the chain length, aggregation only depends on the number of chains in the system. In this study, the system consistently contains two chains. The consistency of the transition temperature  $T_{\text{high}}^*$  across all three PTH systems might indicate that the aggregation process plays a dominant role in the first-order PT, keeping the transition temperature independent of chain length.

Another factor that influences the transition temperature is entropy. When analyzing the temperature dependence of the PT, entropic contributions have to be taken into account. Addi-

tionally, the PRIME20 model uses a mean-field approximation (see section 2.3.1) for its energy potentials. This effectively averages entropy effects into the chain’s energy calculation. Thus, the temperature dependence of the phase transition is governed by a balance between energetic and entropic contributions, i.e. by the free energy  $F = U - TS$ . For the PTH systems, entropy may partially compensate for changes in the energy associated with chain length. For instance, the longer chains potentially exhibit greater configurational entropy, counteracting the energy increase that would otherwise raise the transition temperature. However, these considerations should also hold for homopolymer systems like PolyQ.

In longer chains, higher conformational freedom and consequently greater entropy could arise from the heteropolymeric nature of PTH, where diverse folding and aggregation behaviors along the chain could lead to different entropic contributions to the free energy. This hypothesis has some implications for the structures of PTH: first, it suggests that the energy–entropy balance in the folding and aggregation processes remains consistent across different chain lengths. This potentially indicates a robust folding motif and interaction pattern that is not directly tied to the chain length in the here investigated PTH systems. Second, it implies that shorter chains might adopt more ordered, well-defined structures compared to longer chains. In contrast, the longer chains’ greater conformational freedom should lead to more diverse folded and aggregated structures.

A key difference between the three systems lies in their second-order PT behavior at their respective  $T_{\text{low}}^*$ . While PTH(1-42) and PTH(1-84) exhibit distinct peaks in  $C_V(T)$ , PTH(1-34) shows only a minor shoulder. This suggests that the additional residues in PTH(1-42) and PTH(1-84) may promote further structural rearrangements at lower temperatures. These differences in the PT behavior will be explored in detail in the following sections, focusing on the structural properties of each system. To understand the structural changes accompanying these PTs, observables such as the tensor of gyration and HB maps will be analyzed, as introduced in section 4.2 and analogous to the previous chapters on PolyQ and A $\beta$ . However, before delving into this structural analysis, the next section provides a mapping of reduced model temperatures to physical temperatures, which aids our assessment of the physiological relevance of each system’s phase behavior.

### 6.1.1 Temperature mapping

In order to convert the reduced temperatures in the PRIME20 model to physical temperatures, it is essential to determine the energy scale of the HB interactions  $\varepsilon_{\text{HB}}$ . With knowledge of  $\varepsilon_{\text{HB}}$ , the reduced model temperature  $T$  can be converted to a physical temperature  $T'$  using the relation  $T' = T \varepsilon_{\text{HB}}/k_B$ , as discussed in section 2.3.1.

A good candidate for comparison was the study by Hansmann [214], which investigated the folding behavior of monomeric PTH(1-34) using a multi-canonical sampling approach. The study employed an all-atom model with implicit solvent interactions. With their simulation technique, they were able to calculate the specific heat  $C_V(T)$  over a wide temperature range. The work of Hansmann is particularly relevant to the present study, as the specific heat is a key observable for analyzing PT behavior here as well. Hansmann reported a helix-coil transition at  $T'^* = 560\text{K}$ , marked by a sharp peak in  $C_V(T)$ . By aligning the peak of the folding transition in this study with the peak in  $C_V(T)$  from Hansmann, the temperature conversion factor ( $\varepsilon_{\text{HB}}/k_B$ ) is determined to be 3000K. This results in the following conversion rule for the temperature:

$$T' = T \cdot 3000\text{K}. \quad (6.1)$$

Using this relationship, room temperature in reduced PRIME20 model units corresponds to  $T = 0.100$ , placing room temperature in the low-temperature phase, below the second-order PT signatures observed in all three PTH systems.

A similar study by Kim et al. [215] used the WLMC sampling method, which is closely related to SAMC, to investigate the folding of monomeric PTH(1-34), using an all-atom force

field. In their results, two peaks in the specific heat, were identified: one at  $T_1 = 561\text{K}$  and another at  $T_2 = 424\text{K}$ . Kim et al. attributed the peak at  $T_1$  to the chain's collapse transition and the peak at  $T_2$  to its folding transition. Notably, the temperature  $T_1$  aligns with the results of Hansmann. However, the assignment of these peaks to the folding and collapse transitions in Kim et al.'s study lack structural analysis of phase states. Due to this lack of information, the study can be considered less suitable for comparison to the present work. Instead, it primarily serves as support for the conversion factor  $(\varepsilon_{\text{HB}}/k_B) = 3000\text{K}$  derived from Hansmann.

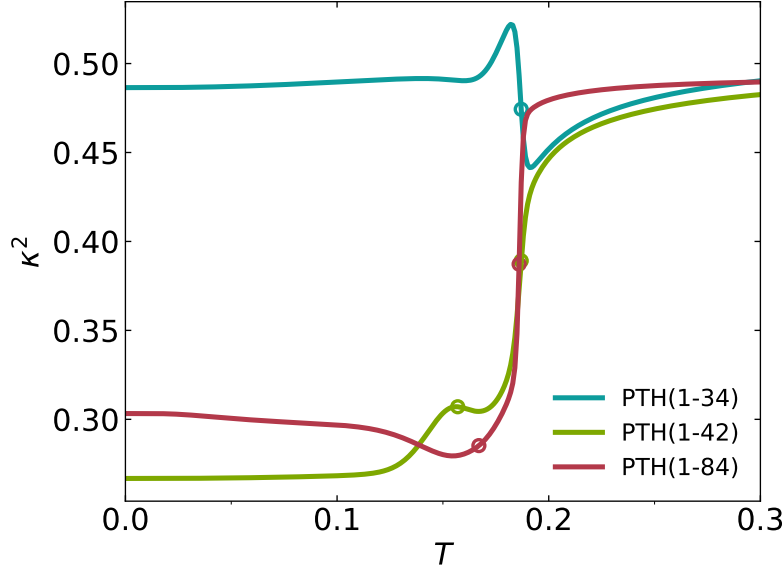
An experimental study by Liu et al. [216] used small-angle neutron scattering (SANS) to determine the radius of gyration  $R_g$  for PTH(1-34) at two different temperatures. The reported  $R_g$  were  $R_g(T = 37^\circ\text{C}) = 12.7\text{\AA}$  and  $R_g(T = 55^\circ\text{C}) = 12.9\text{\AA}$ . When comparing these two  $R_g$  values with the temperature dependent  $R_g$  curve of the PTH(1-34) system simulated in this study (see fig. D.1(b)), one gets two respective conversion factors:  $\varepsilon_{\text{HB}}/k_B = 1802\text{K}$  and  $1874\text{K}$ . This places room temperature at  $T = 0.166$  and  $T = 0.160$  respectively. However, the conversion factors based on the comparison with  $R_g(T)$  are inherently of low precision, since  $R_g(T)$  stabilizes in the low temperature phase. Therefore, even minor fluctuations in the considered  $R_g$  values significantly alter the resulting conversion factor. For instance, if one were to use  $T = 0.100$  from the study of Hansmann, the corresponding  $R_g$  would be  $R_g(T = 0.100) = 12.0\text{\AA}$ , which also would be close to the  $R_g$  values reported by Liu et al.

The conversion factor  $(\varepsilon_{\text{HB}}/k_B) = 3000\text{K}$ , derived from the comparison with the study of Hansmann, is deemed the most reliable for this study and will be applied in the subsequent analyses. However, this conversion factor places the aggregation transition at an unrealistically high temperature, which is likely not physically correct. This conflicts with the observation that aggregates of PTH chains can be dissolved by the human body. A transition temperature of approximately  $560\text{K}$  would make this scenario improbable. To the author's knowledge, no experimental data currently exist to determine a more accurate conversion factor. Consequently,  $(\varepsilon_{\text{HB}}/k_B) = 3000\text{K}$  will be used as a provisional benchmark, with the understanding that it should be revisited and updated when more fitting experimental data becomes available. Importantly, this conversion factor does not impact the structural analysis across the different phases, as it only puts the room temperature on a different position on the temperature-axis without altering the overall shape of the graphs.

## 6.2 Structural analysis

### 6.2.1 Relative shape anisotropy

The tensor of gyration  $\mathbf{T}_g$  is an insightful observable for evaluating the shape of a polymer chain. It was already used in previous chapters of this work to investigate dimer systems of PolyQ and A $\beta$ . As the temperature decreases, all three eigenvalues,  $\lambda_i$  with  $i \in \{1, 2, 3\}$ , of  $\mathbf{T}_g$  exhibit a sharp increase at  $T_{\text{high}}^*$ , revealing a collapse of the chains at the first-order PT temperature. Plots of the  $\lambda_i(T)$  can be found in fig. D.1(a) in appendix D. From the eigenvalues of  $\mathbf{T}_g$ , the relative shape anisotropy  $\kappa^2$  can be calculated, as described by eq. (3.46).  $\kappa^2(T)$  is displayed in fig. 6.3 for all three PTH systems. In this and subsequent figures in this chapter, circles mark the PT temperatures,  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ , identified from the peaks in the canonical heat capacity  $C_V(T)$ . At high temperatures,  $\kappa^2$  is close to 0.5 for all systems, consistent with a random coil conformation typical of the disordered high-temperature phase. As the temperature decreases,  $\kappa^2$  initially declines across all systems. However, as  $T$  approaches  $T_{\text{high}}^*$ , the behavior of the PTH(1-34) system diverges; here,  $\kappa^2$  changes slope and begins to increase, suggesting an elongation of the chain conformations. This increase is only brief, followed by a decline that returns  $\kappa^2$  to values close to 0.5, similar to the high-temperature phase. In contrast, for the other PTH(1-42) and PTH(1-84) systems,  $\kappa^2$  continues to decrease at  $T_{\text{high}}^*$ , and stabilizes at lower values for temperatures  $T < T_{\text{high}}^*$ . This suggests that at  $T_{\text{high}}^*$  the longer chains fold into more



**Figure 6.3:** Relative shape anisotropy  $\kappa^2$  of the three PTH systems. Circles mark the system's PT temperatures.

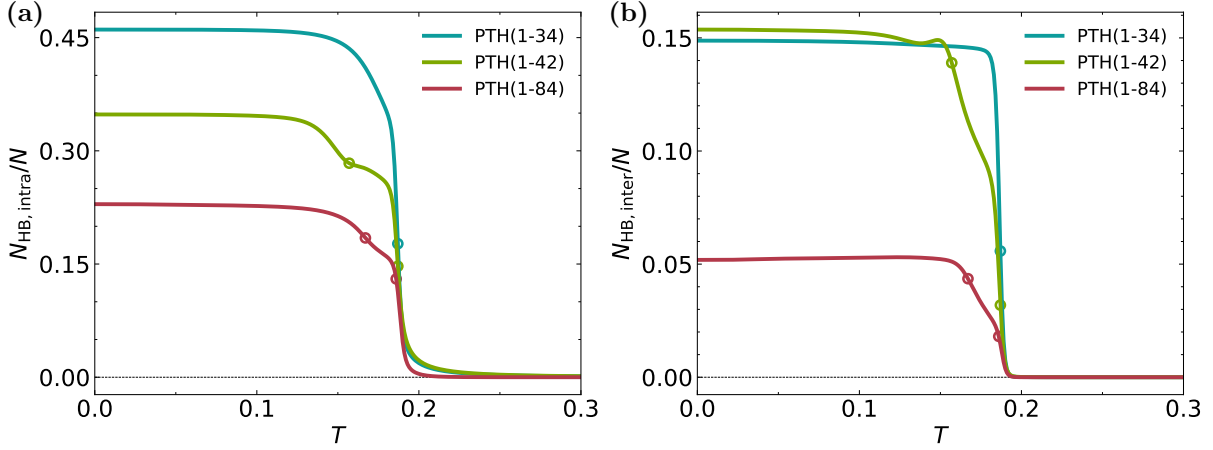
spherical and relatively more compact conformations compared to PTH(1-34). At the secondary PT temperature  $T_{\text{low}}^*$ , minor changes in  $\kappa^2$  are observed for PTH(1-42) and PTH(1-84), hinting at additional structural rearrangements, though these changes are less pronounced than those near  $T_{\text{high}}^*$ .

The behavior of  $\kappa^2$  reveals differences in the shape of the folded structures that the three PTH systems adopt in the low-temperature phase. The PTH(1-34) system exhibits more rigid and extended structure below  $T_{\text{high}}^*$  compared to the more spherical structures of PTH(1-42) and PTH(1-84). A possible explanation would be, that the additional residues in PTH(1-42) and PTH(1-84) are more flexible than the PTH(1-34) fragment. They are therefore able to wrap around themselves more easily, adopting more spherical shapes. By complementing the structured PTH(1-34) fragment, they can ultimately give the entire chain a more spherical configuration. This behavior is consistent with the idea that the structured and more rigid segment of PTH resides in the PTH(1-34) fragment.

## 6.2.2 Intra vs. inter-molecular contacts

Figures 6.4(a) and 6.4(b) show the temperature dependence of the number of established intra- and inter-molecular HBs, respectively. The values are normalized by the total number of amino acid residues in the systems and will be denoted as  $\mathcal{N}_{\text{HB,intra}}(T) = N_{\text{HB,intra}}(T)/N$  and  $\mathcal{N}_{\text{HB,inter}}(T) = N_{\text{HB,inter}}(T)/N$ . At high temperatures, both  $\mathcal{N}_{\text{HB,intra}}$  and  $\mathcal{N}_{\text{HB,inter}}$  are close to zero, which reflects the random coil conformations of the chains, where few, if any, HBs are formed, neither between chains, nor within a single chain. Upon cooling, both  $\mathcal{N}_{\text{HB,intra}}$  and  $\mathcal{N}_{\text{HB,inter}}$  increase at  $T_{\text{high}}^*$ , which indicates that the PT at  $T_{\text{high}}^*$  encapsulates both a folding and an aggregation transitions.

Below  $T_{\text{high}}^*$ , the behavior of  $\mathcal{N}_{\text{HB,intra}}$  diverges among the three PTH systems. Around  $T_{\text{low}}^*$ , PTH(1-42) and PTH(1-84) exhibit a second increase in  $\mathcal{N}_{\text{HB,intra}}$ . PTH(1-34) on the other hand does not show signs of a second-order PT and  $\mathcal{N}_{\text{HB,intra}}$  increases only at  $T_{\text{high}}^*$  with a gradual change in slope. The behavior of PTH(1-42) and PTH(1-84) suggests that the second-order PT in the longer chains involves further energy optimization of the systems by favoring configurations that increase intra-molecular HBs. Another interesting observation is that in the folded state at low temperatures, the PTH(1-34) system has a higher percentage of its beads involved in intra-molecular HBs, compared to the other two systems. PTH(1-84) shows the lowest value for



**Figure 6.4:** Number of intra- and inter-molecular hydrogen bond contacts vs. the temperature for the three PTH systems and normalized by the number of residues in the respective systems. (a) Intra-molecular HB contacts  $N_{\text{HB,intra}}(T)$ . (b) Inter-molecular HB contacts  $N_{\text{HB,inter}}(T)$ .

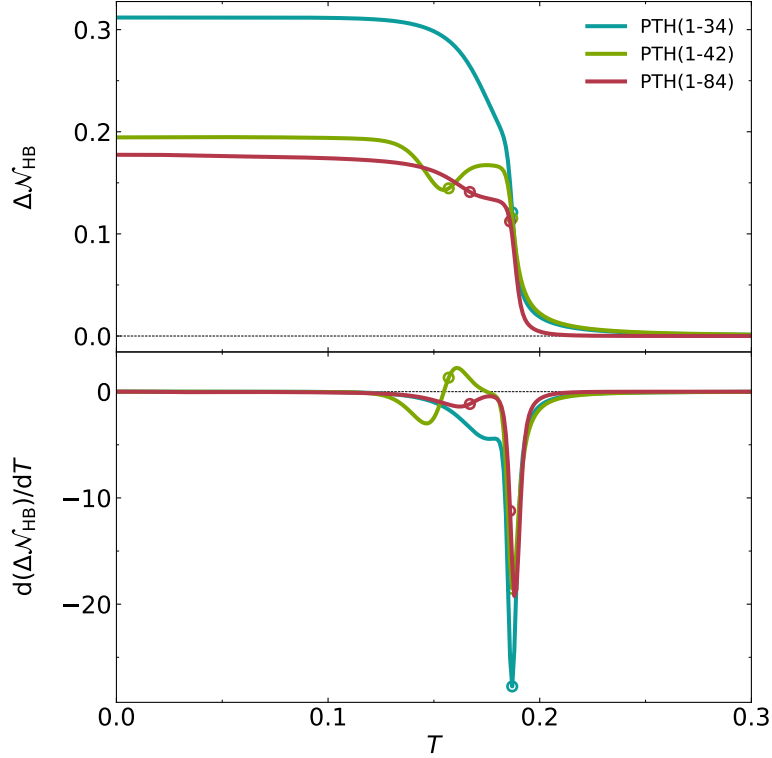
$\mathcal{N}_{\text{HB,intra}}$  of the three. This highlights the importance of the PTH(1-34) fragment to the folding behavior of the PTH chain, as the additional residues in PTH(1-42) and PTH(1-84) contribute less to the formation of intra-molecular HBs.

When looking at the inter-molecular HB contacts, more differences between the three PTH systems become apparent. First, the increase of  $\mathcal{N}_{\text{HB,inter}}$  at  $T_{\text{high}}^*$  is most pronounced for PTH(1-34). When going from the high-temperature phase to the low-temperature one,  $\mathcal{N}_{\text{HB,inter}}$  of PTH(1-34) shows a steep increase after which it quickly flattens and remains at a value around 0.15. In contrast, the two larger PTH systems show a change in slope around  $T_{\text{low}}^*$ , similar to what was already observed for the intra-molecular HBs. Since  $\mathcal{N}_{\text{HB,inter}}$  of PTH(1-42) and PTH(1-84) further increases below  $T_{\text{low}}^*$ , changes in the inter-molecular HB network of the aggregated structures are likely another contributing factor to the second-order PT. Notably, PTH(1-84) has the lowest value of  $\mathcal{N}_{\text{HB,inter}}$  in the low temperature phase, compared to the two shorter systems. Complementary to the observations from intra-molecular HB contacts, this supports the idea that the PTH(1-34) fragment is the key region in driving both the folding and the aggregation transition as the additional residues in PTH(1-42) and PTH(1-84) seem to contribute less to these processes.

In order to compare the relative importance of intra- and inter-molecular HBs in the three PTH systems, the difference  $\Delta\mathcal{N}_{\text{HB}} = \mathcal{N}_{\text{HB,intra}} - \mathcal{N}_{\text{HB,inter}}$  is calculated and displayed in fig. 6.5. Over the entire temperature range,  $\Delta\mathcal{N}_{\text{HB}}$  is positive, indicating that in all three systems the intra-molecular HB contacts are in the majority. When cooling from the high-temperature phase towards  $T_{\text{high}}^*$ ,  $\Delta\mathcal{N}_{\text{HB}}$  increases for all systems, indicating that the folding transition is the dominant process of the combined folding and aggregation transition. Below  $T_{\text{high}}^*$ , PTH(1-34) maintains a significantly higher value of  $\Delta\mathcal{N}_{\text{HB}}$  compared to the two larger systems, reflecting the importance of intra-molecular HBs to its folded state. For the two larger systems, their second-order PT at  $T_{\text{low}}^*$  produces features in  $\Delta\mathcal{N}_{\text{HB}}$  that are not present in PTH(1-34). Specifically, a local minimum can be observed for the PTH(1-42) system. This indicates that the formation of inter-molecular HBs plays a stand-out role in the second-order PT of PTH(1-42). In contrast, PTH(1-84) shows only minor alterations in the slope of  $\Delta\mathcal{N}_{\text{HB}}$  at  $T_{\text{low}}^*$ .

### 6.2.3 Hydrogen bond contact probability maps

In order to identify typically conformations that are adopted by the PTH dimer systems in the different phases, the HB networks at different temperatures are analyzed using HB maps. This method visualizes the probability of HB formation between different amino acid residues and



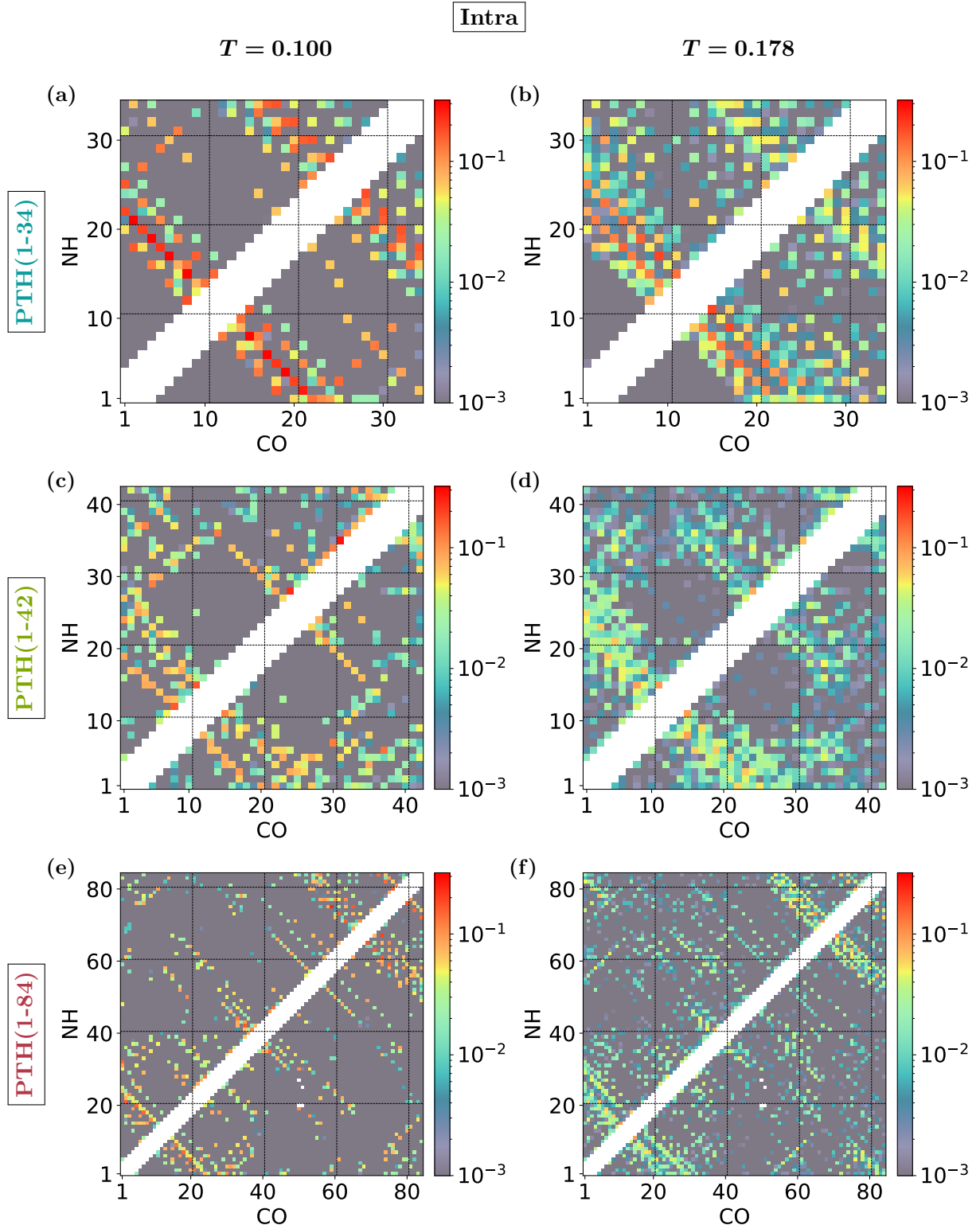
**Figure 6.5:** Difference between intra and inter-molecular hydrogen bond contacts as a function of temperature. The top plot shows the difference  $\Delta\mathcal{N}_{\text{HB}}(T) = \mathcal{N}_{\text{HB,intra}}(T) - \mathcal{N}_{\text{HB,inter}}(T)$ , while the bottom plot shows the derivative of  $\Delta\mathcal{N}_{\text{HB}}(T)$ . Circles mark the transition temperatures  $T^*$  of the PTH systems.

has been introduced and discussed extensively for the PolyQ and  $A\beta$  systems. For PTH, two temperatures are selected for the analysis:  $T = 0.100$  and  $T = 0.178$ . Both represent folded and aggregated phases of the systems. However,  $T = 0.178$  lies just below the first-order PT temperatures  $T_{\text{high}}^*$ , while  $T = 0.100$  is in the low-temperature phase, below the second-order PT temperature  $T_{\text{low}}^*$  found in the PTH(1-42) and PTH(1-84) systems. Furthermore, when applying the temperature mapping derived from comparison with simulation results from Hansmann [214] and Kim et al. [215] (see section 6.1.1), the temperature  $T = 0.100$  corresponds to a physical temperature of  $T' = 300\text{K}$ , which is close to room temperature.

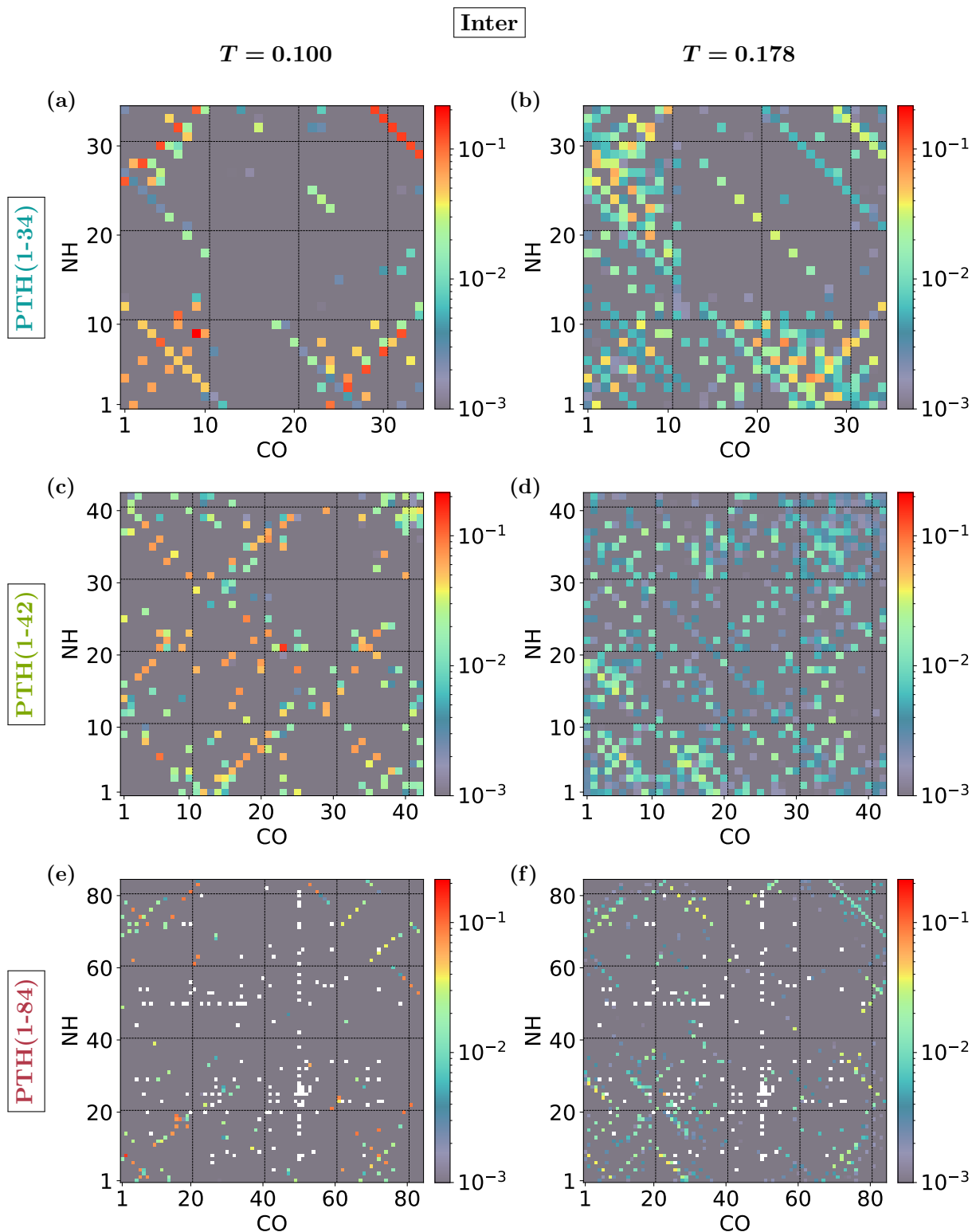
Intra-molecular HB maps are shown in fig. 6.6, while inter-molecular HB maps are displayed in fig. 6.7. Rows in the figures belong to the different PTH systems, while columns correspond to the different temperatures. Equivalent to the representation in the  $A\beta$  chapter, the HB maps use a logarithmic probability scale, where red indicates a high probability of HB formation, while blue indicates a low probability. The scales are displayed to the right of each HB map. Within the same figure, the minimum and maximum values of the scales are kept constant to allow for a direct comparison between the different systems and temperatures. However, the scales differ between the intra- and inter-molecular HB maps. This increases readability of the inter-molecular HB maps since the probability of HB formation is generally higher for intra-molecular HBs, as discussed in the previous section.

### Hydrogen bond networks at $T = 0.178$

The right column of fig. 6.6 shows the intra-molecular HB maps of the PTH systems at  $T = 0.178$ . They reveal distinct regions of high HB probability for each of the three different PTH systems. The intra-molecular HB map of PTH(1-34) exhibits high-probability signatures of descending slope. They stretch out from the turn region, in which no HBs can form and thus leads



**Figure 6.6:** Intra-molecular hydrogen bond contact probability maps for PTH(1-34), PTH(1-42) and PTH(1-84). The probability color scale is logarithmic. The first row displays the HBmaps for PTH(1-34), the second row for PTH(1-42) and the third row for PTH(1-84). The left column shows HBmaps for  $T = 0.100$  and the right column for  $T = 0.178$ .



**Figure 6.7:** Inter-molecular hydrogen bond contact probability maps for PTH(1-34), PTH(1-42) and PTH(1-84). The probability color scale is logarithmic. The first row displays the HBmaps for PTH(1-34), the second row for PTH(1-42) and the third row for PTH(1-84). The left column shows HBmaps for  $T = 0.100$  and the right column for  $T = 0.178$ .

to the white area on the main diagonal. The high-probability signatures are clear indications of  $\beta$ -hairpin structures. Most noticeable are the two  $\beta$ -hairpins at the N-terminal region, around the turn contacts (8,14) and (11,15). Several other  $\beta$ -hairpin signatures of lower probability can be found in adjacent residues. Another region that shows increased  $\beta$ -hairpin probability lies at the C-terminal end of PTH(1-34), close to the turn contacts (22,28) and (24,28). Apart from  $\beta$ -hairpins, there are other antiparallel signatures present in the HB maps of PTH(1-34), that are not close to the turn region. Such structures hint at the complexity of the chain configurations, where chain segments which are far apart along the polymer sequence, are able to form HBs resulting in antiparallel  $\beta$ -sheets.

A configuration fitting the two observed  $\beta$ -hairpin areas is an S-shaped folded structure. Indeed, such configurations can frequently be found in simulation snapshots of PTH(1-34). Examples are shown in figs. 6.8(a), 6.8(b) and 6.8(d). An example of intra-molecular  $\beta$ -sheets between chain segments that are far apart in the sequence is shown in fig. 6.8(d) which features a large loop in one chain.

The PTH(1-42) and PTH(1-84) display similar behavior to PTH(1-34) in their HB maps at  $T = 0.178$ . The  $\beta$ -hairpin structures observed in the PTH(1-34) system also appear in the same regions within these longer chains. However, while these  $\beta$ -hairpins remain the most probable signatures within their respective HB maps, their probabilities are notably lower than those in the PTH(1-34) system. This reduction in probability is accompanied by a broader distribution of probable HB partners for each residue. Regions of relatively high probability are more expansive, with probabilities more evenly spread across these areas. This distribution suggests that the additional residues in PTH(1-42) and PTH(1-84) allow for a wider range of conformations. This effect is especially pronounced in the PTH(1-84) system, where the HB maps indicate multiple likely HB partner per residue, indicating a diverse set of possible conformations.

In addition to the  $\beta$ -hairpin structures at the N-terminal end present in all three PTH systems, the HB map of PTH(1-84) reveals additional  $\beta$ -hairpin formations involving residues of the additional chain segment PTH(35-84). Notable examples include  $\beta$ -hairpins around the turn contacts (37,41), (56,60), and (76,80). Particularly prominent are several  $\beta$ -hairpin signatures of high probability, with the most distinct ones involving turn contacts (64,68), (65,69), (66,70), and (67,71).

As with the PTH(1-34) system, the PTH(1-42) and PTH(1-84) systems exhibit antiparallel  $\beta$ -sheet signatures that do not originate directly from the turn region. These signatures indicate complex chain conformations in the system, that allow distant chain segments to come into contacts and form HBs. In PTH(1-84), for instance, this creates the anti-parallel  $\beta$ -sheet [(1,81) – (3,80) – (5,78) – (7,76)] as evident from fig. 6.6(f). Besides anti-parallel  $\beta$ -sheets, signatures of parallel  $\beta$ -sheets also appear in both the PTH(1-42) and PTH(1-84) systems. They manifest as diagonal structures of high HB probabilities with an ascending slope. In the HB map of PTH(1-84), for example, residues [(2,56) – (4,58) – (6,60) – (8,62) – (10,64) – (12,66)] show elevated contact probabilities, indicating the formation of a parallel  $\beta$ -sheet (for an example configuration see fig. 6.8(i)). In the PTH(1-42) system, the presence of residues beyond the 34th position also facilitates parallel  $\beta$ -sheet formation. The absence of this feature in the shorter PTH(1-34) system suggests that the additional residues in PTH(1-42) and PTH(1-84) contribute to the formation of parallel  $\beta$ -sheets.

Notably,  $\alpha$ -helix signatures are absent in the HB maps of all PTH systems at  $T = 0.178$ . This absence is somewhat unexpected, a topic which will be further explored towards the end of this section on the analysis of HB maps.

The inter-molecular HB maps of the PTH systems at  $T = 0.178$  are displayed in the right column of fig. 6.7. In the HB map of the PTH(1-34) system, signature of both parallel and anti-parallel  $\beta$ -sheets are visible. The most probable  $\beta$ -sheet structures involve beads at the N-terminal end of the chain.  $\beta$ -sheet signatures in which N-terminal beads of both chains participate show a preferred parallel alignment (see for example fig. 6.8(d)) Notable patterns

include  $[(N_1, C_2) - (N_3, C_4) - (N_5, C_6) - (N_7, C_8)]$  and  $[(N_4, C_1) - (N_6, C_3) - (N_8, C_5) - (N_{10}, C_7)]$ . By contrast,  $\beta$ -sheet signatures linking N-terminal residues of one chain to C-terminal residues of the other show both parallel and anti-parallel orientations. Together with the intra-molecular  $\beta$ -hairpins and  $\beta$ -sheets identified, these inter-molecular HBs create a network of  $\beta$ -sheets spanning both chains. This leads to multi-layered  $\beta$ -sheet configurations. Examples of such configurations are shown in figs. 6.8(a) to 6.8(d).

In the PTH(1-42) systems, maximal contact probabilities in the HB map are reduced compared to those of PTH(1-34). As was explored in section 6.2.2, PTH(1-42) has fewer established HBs at  $T = 0.178$  compared to PTH(1-34). Nevertheless, both parallel and anti-parallel  $\beta$ -sheet signatures remain visible in fig. 6.7(d). However, these signatures differ in some areas from the ones found in the HB map of PTH(1-34). For instance, parallel and anti-parallel  $\beta$ -sheet signatures connecting residues [1–10] of one chain with residues [18–34] of the other chain are largely absent in the PTH(1-42) system. Example configurations of the PTH(1-42) system forming  $\beta$ -hairpins and parallel, as well as anti-parallel  $\beta$ -sheets, are shown in figs. 6.8(e) to 6.8(g).

The inter-molecular HB map of PTH(1-84) reveals prominent signatures of both parallel and anti-parallel  $\beta$ -sheets across different regions of the chain. Some of these patterns resemble those in PTH(1-34), while others involve beads unique to the PTH(35-84) segment (see for example fig. 6.8(i) vs. fig. 6.8(j)). Notably, some cells of the HB map remain empty and white, indicating no HB contacts were observed between the respective beads during the simulation runs. For example, the CO bead of Arg50 residue shows several empty cells. Even with extended simulation times, these empty cells are expected to display a very low probability of HB formation, which suggests limited inter-molecular interactions in this region.

### Hydrogen bond networks at $T = 0.100$

Lowering the temperature to  $T = 0.100$  results in a significant increase in the total probability of intra-molecular HB formation across all three PTH systems, as evident in the HB maps in the left column in fig. 6.6. This observation is consistent with the previous analyses of  $\mathcal{N}_{\text{intra}}$  and  $\mathcal{N}_{\text{inter}}$  discussed in section 6.2.2. Notably, the temperature dependence of  $\mathcal{N}_{\text{intra}}$  (fig. 6.4(a)) correlates with the observed changes in the intra-molecular HB maps when the temperature is reduced to  $T = 0.100$ .

For the three investigated PTH systems, the existing  $\beta$ -hairpin and  $\beta$ -sheet patterns at  $T = 0.178$  not only increase in probability but also become sharper and more distinct, as adjacent low-probability features fade. This temperature reduction, however, does not fundamentally change the intra-molecular structure across the PTH systems. An exception is observed in the PTH(1-42) system, where an  $\alpha$ -helical structure emerges in the C-terminal half of the chain. This  $\alpha$ -helix pattern is characterized by contacts at  $(C_i, N_{i+4})$  for  $i \in \{27, 28, 29, \dots, 35\}$ . Snapshots from the SAMC simulations of the PTH(1-42) system, shown in figs. 6.8(g) and 6.8(h), also display  $\alpha$ -helices towards the C-terminal end, colored in orange. The PTH(1-84) system occasionally exhibits  $\alpha$ -helices in the C-terminal region as well. An example appears in fig. 6.8(k), where an  $\alpha$ -helix forms at the C-terminal end of one chain, with beads colored in red. However, as evident from the HB maps, the probability of forming such  $\alpha$ -helices is significantly lower compared to PTH(1-42).

In contrast, the inter-molecular HB networks of the PTH systems show more drastic changes compared to their intra-molecular HB networks. The inter-molecular HB maps at  $T = 0.100$  are displayed in the left column in fig. 6.7. In the PTH(1-34) system, the diversity of probable inter-molecular HB configurations decreases significantly. The most common arrangements now favor parallel  $\beta$ -sheets, while several of the anti-parallel  $\beta$ -sheet patterns present at  $T = 0.178$  are no longer observed. Notably, the anti-parallel  $\beta$ -sheets between N-terminal residues of one chain and C-terminal residues of the other lose their prominence.

By comparison, the PTH(1-42) and PTH(1-84) systems exhibit a more substantial overall increase in the probability of inter-molecular HB contacts, which aligns with the findings in sec-

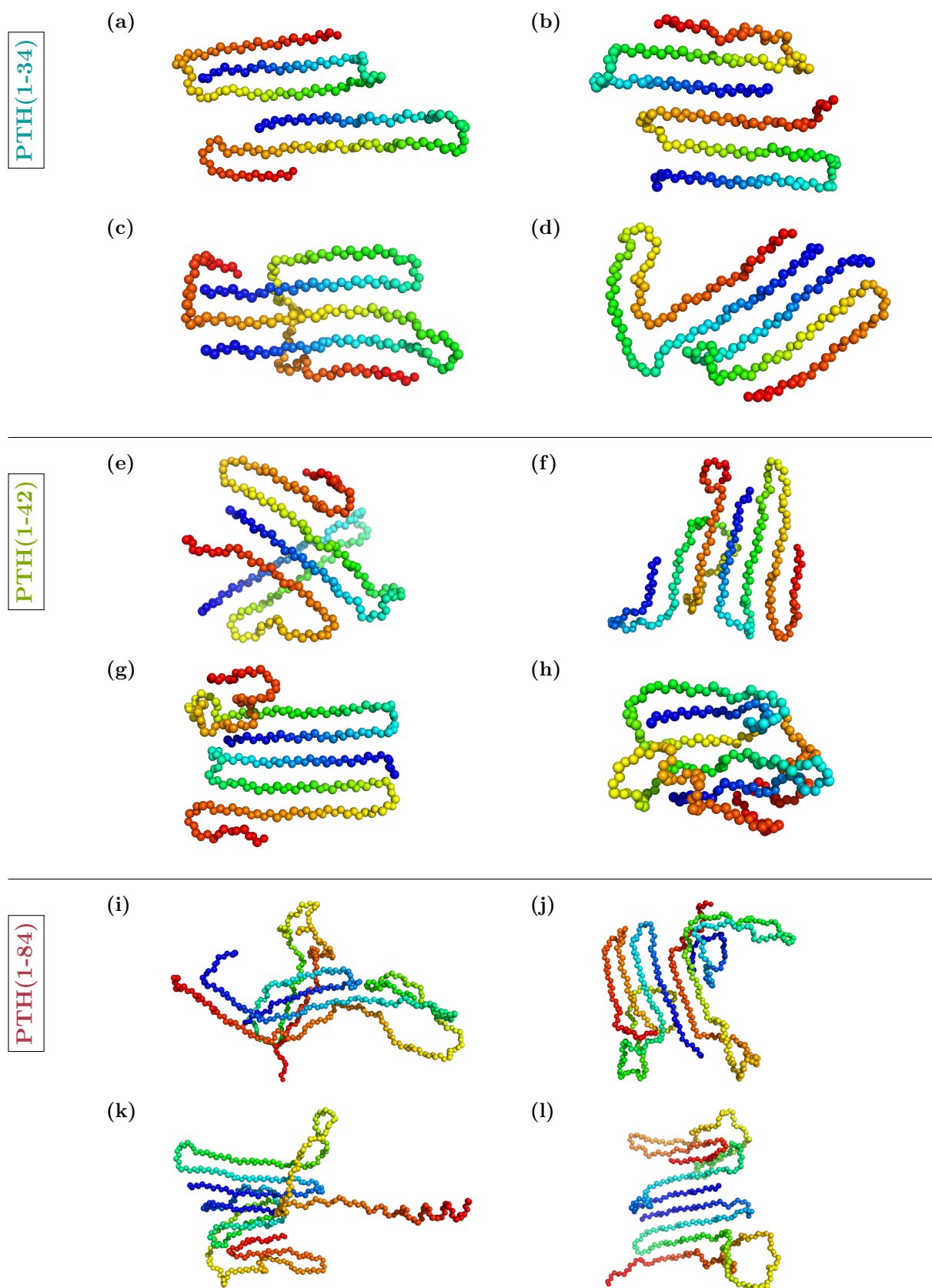
tion 6.2.2. Furthermore, the inter-molecular HB maps patterns differ significantly between the PTH(1-34) system and the two larger systems. While the general trend to form  $\beta$ -sheets persists between the systems, their specific locations and HB partners differ. The PTH(1-42) and PTH(1-84) systems display both parallel and anti-parallel  $\beta$ -sheets. Their aggregation driven second-order PT at  $T_{\text{low}}^*$  leads to distinct aggregation interfaces in their dimer structures compared to the PTH(1-34) system. A shared motif between the three PTH systems remains the formation of parallel and anti-parallel inter-molecular  $\beta$ -sheets. However, the specific configuration of these  $\beta$ -sheets vary. Example of both parallel and anti-parallel inter-molecular  $\beta$ -sheets in PTH(1-42) and PTH(1-84) are shown in figs. 6.8(f), 6.8(g), 6.8(j) and 6.8(l).

In conclusion, the HB maps demonstrate a strong tendency for the PTH dimer systems to form  $\beta$ -sheets when the systems are below the folding and aggregation transition temperature  $T_{\text{high}}^*$ . Intra-molecularly, this leads predominantly to  $\beta$ -hairpins positioned at various but distinct sites along the chain. Notably, recurring  $\beta$ -sheet motifs with turn contacts at or adjacent to (11,14) and (24,28) appear consistently across all three PTH systems. This common structural pattern aligns with the thermodynamic analysis, supporting the hypothesis of robust folding motifs and interaction patterns that remain largely independent of chain length.

Furthermore, intra-molecular  $\beta$ -sheets also form between chain segments that are distant in the protein sequence. This suggests that more complex structures, like for instance configurations featuring large loops, are present within the ensemble of possible PTH dimer configurations. Examples for such large-loop configurations are shown in figs. 6.8(d) and 6.8(f). The diversity in the conformational space of the PTH dimer systems increases with the chain length. This is reflected in the broader distribution of potential HB partners for each residue seen in the HB maps of PTH(1-42) and PTH(1-84) compared to PTH(1-34). The shorter PTH(1-34) exhibits a more focused distribution of probable HB partners, resulting in well-defined  $\beta$ -hairpin structures, and also well-ordered aggregates at  $T_{\text{high}}^*$ . In contrast, the longer PTH(1-42) and PTH(1-84) chains show a broader range of probable HB partners, leading to an increased diversity in the position and extension of  $\beta$ -hairpins and  $\beta$ -sheets. This structural diversity in the folded structures also translates into greater variety of aggregate structures at  $T_{\text{high}}^*$ , allowing for multiple options of energy optimization in the aggregation process.

At lower temperatures, the overall probability of  $\beta$ -sheet formation increases, accompanied by a reduction in the diversity of possible HB partners. In the intra-molecular HB networks,  $\beta$ -hairpins remain the dominant secondary structures, while in the inter-molecular HB networks, more substantial changes can be observed. All three PTH systems show an increase in the probability of forming parallel inter-molecular  $\beta$ -sheets. In PTH(1-34), parallel  $\beta$ -sheet patterns seen at higher temperatures become reinforced, while anti-parallel  $\beta$ -sheet signatures are mostly abandoned. As the PTH(1-42) and PTH(1-84) systems undergo their second-order PT at  $T_{\text{low}}^*$ , which is dominated by the aggregation process (see section 6.2.2), new and distinct parallel and anti-parallel  $\beta$ -sheet patterns emerge. This is accompanied by a significant reduction in the diversity of aggregate structures. The specific positions of the  $\beta$ -sheets vary among the PTH systems, suggesting unique aggregation interfaces for each system. Additionally, in the PTH(1-42) system,  $\alpha$ -helical patterns emerge at the C-terminal end of the chain. This feature is not equally observed in the other PTH systems.

The absence of  $\alpha$ -helices and the dominance of  $\beta$ -sheets in the PTH dimer systems contrasts with studies on the PTH monomer. Experimental research from the 1990s, such as studies by Barden et al. [217], Marx et al. [218], and Willis et al. [219], consistently identified  $\alpha$ -helices as the main structural feature of the PTH monomer in aqueous solution. Computational studies, including Hansmann [214] using multi-canonical sampling, have confirmed this  $\alpha$ -helical tendency. The absence of  $\alpha$ -helices in the PTH dimer systems of this study suggests that dimerization alters the protein's secondary structure significantly. Rather than  $\alpha$ -helices, the PTH dimer systems exhibit a high probability of  $\beta$ -sheet formation. This shift is likely due to



**Figure 6.8:** Configuration snapshots from simulations of the PTH systems: (a)–(d) PTH(1-34), (e)–(h) PTH(1-42), (i)–(l) PTH(1-84). Only the backbone beads are shown: NH, C $_{\alpha}$  and CO. Side chain beads (R) are omitted for better visibility. The beads in each chain are colored using a rainbow gradient, with the N-terminus in blue and the C-terminus in red. The configurations are selected at energies corresponding to temperatures in the folded and aggregated state.

the competition between intra- and inter-molecular HBs introduced by the second PTH chain. Thus, both aggregation and folding transition occur simultaneously at the same PT temperature. The additional network of inter-molecular HBs competes with the intra-molecular interactions, driving the system towards  $\beta$ -sheets rather than the  $\alpha$ -helices, that were observed for PTH monomers.

However, an exception in this trend is the PTH(1-42) system, where an  $\alpha$ -helical pattern appears at the C-terminal end of the chain below  $T_{\text{low}}^*$ , even in the here investigated dimer systems. The PTH(1-84) system occasionally shows similar  $\alpha$ -helical structures (see fig. 6.8(k)). These isolated  $\alpha$ -helices suggest some residual structural preferences in the longer chains, even within the dimer system.

The presence of  $\beta$ -sheets in the PTH dimer systems aligns with the structural characteristics of amyloid fibrils formed by PTH. NMR studies by Gopalswamy et al. have demonstrated a cross- $\beta$  structure<sup>(9)</sup> in the amyloid fibrils of PTH(1-84) [79]. The fibril core of these amyloid fibrils is located around residues 25-37. Within these fibrils, the  $\beta$ -sheets are predominantly aligned in a parallel configuration [220]. Similarly, the PTH dimer systems studied here exhibit a strong tendency for  $\beta$ -sheet formation. Near physiological temperatures, the intra-molecular  $\beta$ -sheets within the PTH dimers also exhibit a preference for parallel alignment, though the participating segments differ from those forming the fibril core in the amyloid structure. Notably, the segment around residues 25-37, which is critical to amyloid fibril formation, shows a strong propensity for intra-molecular  $\beta$ -sheets, particularly  $\beta$ -hairpins, in the PTH dimer systems. This observation suggests that the dimer structures observed here may not directly correspond to immediate precursors to amyloid fibrils. Nonetheless, the marked preference for  $\beta$ -sheet formation in the PTH dimers indicates that dimerization could be an important early step in the pathway towards amyloid fibril development.

## 6.3 Conclusions

The folding and aggregation behavior of PTH dimer systems was analyzed using SAMC simulations in combination with the PRIME20 protein model. Three different PTH systems were investigated: PTH(1-34), PTH(1-42), and PTH(1-84). By analysis of the density of states  $g(U)$  which is the primary result of the SAMC simulations, the distinct phase behavior of the PTH systems was revealed. All three systems exhibit a first-order PT at  $T_{\text{high}}^*$ . Additionally, the PTH(1-42) and PTH(1-84) systems display a second-order PT at lower temperatures. The PT temperatures were determined both from a microcanonical analysis using derivatives of the entropy  $S(U) = \ln g(U)$ , and from a canonical analysis using the specific heat  $C_V(T)$ . A complete list of the determined PT temperatures of the three PTH systems is shown in table 6.1. Through analysis of the number of intra- and inter-molecular HB contacts, the folding and aggregation transitions were found to occur simultaneously at the same PT temperature  $T_{\text{high}}^*$ .

A notable observation is that the temperatures of the PTs at  $T_{\text{high}}^*$  are equal for all three PTH systems, regardless of their differences in chain length. This is unexpected, as folding transitions typically scale with chain length, as demonstrated, for example, in the PolyQ systems. However, unlike PolyQ, the PTH systems are heteropolymers, composed of different types of amino acids. This heterogeneity likely results in distinct folding behaviors of different segments along the chain, which could influence the transition temperature. Additionally, increased entropic contributions to the folding and aggregation transitions of the longer systems may offset the energetic differences, stabilizing the transition temperature. Another crucial factor is the simultaneous folding and aggregation observed in these systems. Unlike folding, aggregation is independent of chain length. This suggests that the aggregation process exerts a dominant influence on the phase transition behavior, effectively overriding the chain length dependence typically associated with folding transitions.

<sup>(9)</sup>For a detailed description of the cross- $\beta$ -structure, see section 2.1.1.

In order to identify structural properties of the different phases in the PTH systems, the intra- and inter-molecular HB networks were analyzed using HB maps. At  $T_{\text{high}}^*$ , the PTH systems exhibit a strong tendency to form  $\beta$ -sheets. The intra-molecular HB maps reveal  $\beta$ -hairpins and  $\beta$ -sheets as the dominant secondary structures in the folded chains. In addition, PTH(1-34) exhibits distinct, well-defined inter-molecular  $\beta$ -sheets both in parallel and anti-parallel alignment. In contrast, the PTH(1-42) and PTH(1-84) systems show a broader range of probable aggregate structures and conformations. However,  $\beta$ -sheet conformations remain prominent. When the temperature is lowered to  $T_{\text{low}}^*$ , the PTH(1-42) and PTH(1-84) systems exhibit a significant increase in the probability of inter-molecular HB contacts. This is accompanied by the formation of distinct parallel and anti-parallel inter-molecular  $\beta$ -sheet patterns. The HB maps further highlight how the conformational space of the PTH dimer systems expands with chain length, as the longer chains exhibit a wider distribution of probable HB partners per residue, enabling a more diverse set of possible conformations and greater entropic freedom.

The  $\beta$ -sheet structures observed in the PTH dimer systems share structural similarities with amyloid fibrils formed by PTH. However, the dimer structures in this study do not directly correspond to immediate precursors of amyloid fibrils, as their specific HB networks differ. Nonetheless, the pronounced preference for  $\beta$ -sheet formation in the PTH dimer systems suggests that dimerization could represent a critical early step in the pathway towards amyloid fibril development.

Interestingly,  $\alpha$ -helices are almost entirely absent in the PTH dimer system's folded phases below  $T_{\text{high}}^*$  and  $T_{\text{low}}^*$ . This contrasts with experimental studies on the PTH monomer, where  $\alpha$ -helices were consistently identified as the predominant structural feature of the (1-34) fragment, while the 50 C-terminal residues are intrinsically disordered [79, 221]. The absence of  $\alpha$ -helices in the PTH dimer systems suggests that dimerization significantly alters the protein's secondary structure. The introduction of inter-molecular HB competition, alongside the intra-molecular HB network, appears to shift the structural preference towards  $\beta$ -sheets rather than the  $\alpha$ -helices observed for PTH monomers. However, isolated  $\alpha$ -helices persist in the PTH(1-42) and PTH(1-84) systems, reflecting residual structural preferences of  $\alpha$ -helix formation in the longer chains even within the dimer system.

## Chapter 7

# Summary and Conclusions

In the here presented work, the dimerization behavior of three different amyloidogenic proteins has been investigated: PolyQ, A $\beta$  and PTH. All of these proteins are known to form amyloid fibrils *in vivo*. The study utilized computer simulations using the PRIME20 protein model. The PRIME20 model is an intermediate-resolution protein model, which enables a protein-sequence-specific representation of peptides by considering different interaction parameters between the different naturally occurring amino acids. Furthermore, it provides detailed structural insights, such as backbone dihedral angles and directional backbone HBs. The SAMC algorithm is used for sampling of configuration space. SAMC is an advanced flat-histogram MC method, that achieves an even sampling of the energy space, a feature particularly advantageous for thermodynamic and structural investigations of complex systems like proteins. Additionally, SAMC provides an approximation of the mDOS, a quantity from which the entire thermodynamics of the system can be derived.

The three proteins examined in this study were selected for their distinct fibril state properties. While all share the characteristic amyloid fibril structural motif, their stability and physiological implications differ significantly. In PolyQ and A $\beta$ , the amyloid state represents a pathological denatured form which is closely associated with Huntington's and Alzheimer's disease, respectively. In contrast, fibrils of PTH represent a functional state, which the body can dissolve as required. Understanding the thermodynamic properties governing the early aggregation stages of aggregation in these proteins is essential for unraveling the molecular mechanisms of amyloid formation. This study marks an important step in towards this goal. Such insights are pivotal for advancing our understanding of amyloid-related diseases and developing targeted therapeutic strategies.

### 7.1 Summary of results

Dimer systems of PolyQ chains of various chain lengths, ranging from 14 to 36 residues, were simulated to investigate their thermodynamic and structural behavior. The thermodynamic analysis revealed that all PolyQ systems undergo folding and aggregation transitions. For each system, the folding and the aggregation transition occur simultaneous on the temperature scale. Furthermore, the phase transition temperature,  $T^*$ , of this simultaneous transition was found to be chain length dependent. For short chains ( $N_{\text{res}} < 22$ ),  $T^*$  lies below room temperature, whereas for long chains ( $N_{\text{res}} \geq 22$ ),  $T^*$  lies above room temperature. The chain-length dependence of the combined folding and aggregation transitions was shown to scale with  $N^{-1/3}$ . The temperature dependence is consistent with previous experimental observations by Walters et al. [26] who showed that PolyQ chains are in the collapsed state at room temperature.

Structural analysis of the aggregated state revealed that all investigated PolyQ dimer systems form  $\beta$ -hairpins and intermolecular  $\beta$ -sheets. Interestingly, short-chain systems form intermolecular  $\beta$ -sheets composed of entire chains as well as aggregated  $\beta$ -hairpins, while long-chain

systems only form the latter.

Further examination of intra- and inter-molecular HBs highlighted distinct mechanisms driving the PT for short and long chains. For shorter chains ( $N_{\text{res}} \leq 24$ ), the PT at  $T^*$  is driven predominantly by the formation of intermolecular HBs, with aggregation as the dominant process. In contrast, longer chains ( $N_{\text{res}} > 24$ ) undergo a PT dominated by intramolecular HBs, indicating that folding is the key process at the transition. These findings are consistent with experimental studies identifying  $N_{\text{res}} = 24$  as the critical chain length for single PolyQ chain to acts as a nucleus for PolyQ aggregation [26, 36, 37, 164].

The results of the PolyQ investigations have been published in [56].

The dimerization behavior of  $A\beta$  was investigated for its two variants,  $A\beta(1-40)$  and  $A\beta(1-42)$ . Despite their strong similarities in their primary sequences, the two isoforms exhibit distinct phase behavior. Thermodynamic analysis revealed two PTs for both systems, occurring at  $T_{\text{low}}^*$  and  $T_{\text{high}}^*$ . Analysis of temperature-dependent intra- and inter-molecular HB formation identified the first-order PT at  $T_{\text{high}}^*$  as a folding and aggregation transition. In contrast, the second-order PT at  $T_{\text{low}}^*$  was linked to rearrangements of the HB networks within the aggregated state. Mapping these transition temperatures to physical units revealed that room temperature lies between  $T_{\text{low}}^*$  and  $T_{\text{high}}^*$ . This behavior is consistent with the aggregation-prone nature of  $A\beta$ . In the thermodynamic analysis of the  $A\beta$  systems it already becomes apparent that the  $A\beta(1-40)$  and  $A\beta(1-42)$  systems exhibit distinct aggregation behavior, as the PT temperatures of the two systems differ significantly. These differences became even more apparent in the structural analysis of the aggregated state.

Identification of the secondary structure elements in the aggregated state was achieved by analysis of HB maps. A common structural motif between both  $A\beta$  systems were  $\beta$ -hairpins in the N-terminal region and around the turn contact (Val24-Lys28). Strong polymorphism was observed in the aggregated structures of both  $A\beta$  systems, consistent with previous experimental and computational studies [173, 222, 223]. Both parallel and antiparallel alignment of inter-molecular  $\beta$ -sheets were present. However, significant differences emerged around room temperature, where the  $A\beta(1-42)$  system exhibited a well-defined aggregate structure involving the middle region of the chain. In contrast, the  $A\beta(1-40)$  systems showed greater structural diversity and less pronounced inter-molecular order. This observation aligns with the faster aggregation kinetics of  $A\beta(1-42)$  compared to  $A\beta(1-40)$  [24].

Finally, analysis derived from the systems' Ramachandran plots revealed that Gly residues in the C-terminal region acts as a secondary structure breaker, preventing the formation of a regular secondary structure in this region.

The final protein system investigated in this study was PTH. Three different lengths were considered: PTH(1-34), PTH(1-42), and PTH(1-84). Thermodynamic analysis revealed that all three PTH systems exhibit a first-order PT at  $T_{\text{high}}^*$ . This transition again corresponds to simultaneous folding and aggregation, as revealed by the comparison of intra- and inter-molecular HBs. Additionally, the PTH(1-42) and PTH(1-84) systems displayed a second-order PT at lower temperatures.

A notable and unexpected observation was that the first-order PT occurs at approximately the same  $T_{\text{high}}^*$  for all three PTH systems. This behavior was discussed in the context of heteropolymeric systems and entropy-driven phase transitions. The simultaneity of the folding and aggregation processes was also identified as a key factor contributing to the observed behavior.

Structural analysis of the PTH systems' state below  $T_{\text{high}}^*$  revealed that the aggregated state consists primarily of  $\beta$ -hairpins and  $\beta$ -sheets, similar to the structures observed in the two previous systems. However, chain length influenced the diversity of aggregated structures: the systems of longer chain exhibited a broader distribution of probable HB networks, resulting in greater structural polymorphism. However,  $\beta$ -sheets remained the dominant secondary structure

motif. When decreasing the temperature towards  $T_{\text{low}}^*$ , the PTH(1-42) and PTH(1-84) systems showed a pronounced increase in inter-molecular HBs, forming distinct, well-defined parallel and antiparallel  $\beta$ -sheets. In contrast, the shorter PTH(1-34) system exhibited well-defined aggregate structures already below  $T_{\text{high}}^*$ , which became increasingly refined with further cooling. The resulting aggregate structures of the PTH systems share similarities with the fibril structures of PTH.

Interestingly,  $\alpha$ -helices were rarely observed in the PTH dimer systems, even though they are a prominent feature of the monomer solution structure [79, 221]. This suggests, that the presence of the second chain in the systems significantly alters the preferred secondary structure of the individual chains. The competition between intra- and inter-molecular HBs operating on the same energy-scale was discussed as a likely explanation for this behavior.

## 7.2 Comparison of Amyloid Beta and Parathyroid Hormone

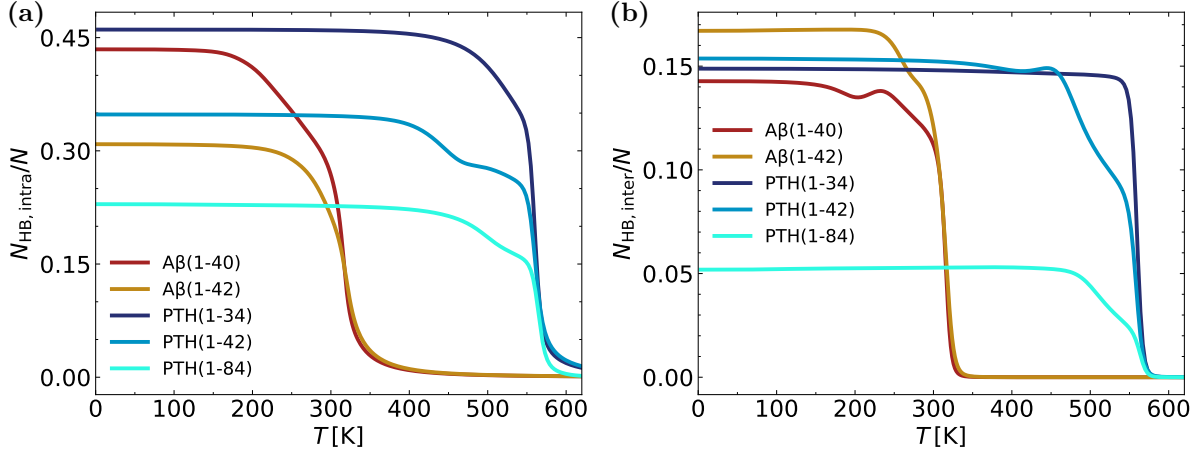
The PTH's ability to form amyloid fibrils is of particular interest when comparing it to  $A\beta$ . Both proteins can be found *in vivo* in an amyloid state with a cross- $\beta$  structure motif [78, 79]. However, not only does the presence of amyloid fibrils have wildly different pathological implications, but the human body shows different capabilities of dissolving these fibrils. While the amyloid state of  $A\beta$  is associated with Alzheimer's disease and the body has a hard time dissolving these fibrils, the amyloid state of PTH is suggested to be a storage form of the protein [44] which can be dissolved by the body. Therefore, a comparison of the results for the dimerization behavior of the PTH and  $A\beta$  systems can provide valuable insights into the differences in their fibrillation behavior and the stability of their amyloid fibrils.

The different stabilities of amyloid fibrils formed by PTH and  $A\beta$  in the human body suggest notable differences in the thermodynamic properties of these proteins. Given that amyloid fibrils of PTH are less stable than those of  $A\beta$ , one might reasonably expect the aggregation temperature of PTH to be lower than that of  $A\beta$ . However, the opposite is observed. From the analysis of the  $A\beta$  systems, the aggregation temperature of  $A\beta$  is approximately 320K, while for PTH, it is around 560K. This apparent contradiction is counterintuitive, as the lower stability of PTH fibrils would suggest a lower aggregation temperature, closer to physiological temperature.

Several factors help explain this inconsistency. First, it is important to note that the analyzed process here involves dimerization rather than the formation of fibrils. Therefore, no direct conclusions about fibril stability can be drawn from the here presented results. Furthermore, experimental studies by Gopalswamy et al. [79] identified specific solvent conditions, including a high pH of 9.0 and a high temperature of 65°C, as suitable conditions for PTH fibril formation. The PRIME20 model, which employs a mean-field approximation to calculate effective interaction potentials, may not fully capture the solvent quality dependencies. Consequently, the physical temperature scale derives in this study could be subject to further refinement.

Additionally, the interaction parameters of PRIME20 are derived from known protein structures from the PDB. At the elevated temperatures around 560K, far above the conditions the parameters were originally optimized at, these force field parameters might lack the accuracy needed to describe the system correctly. This limitation also applies to the simulation results of Hansmann [214], which were used to establish the temperature conversion from reduced PRIME20 units to Kelvin. The force fields employed in those simulations likely encounter similar constraints at high temperatures, potentially affecting the interpretation of the temperature scales and aggregation behaviors in both studies.

Another way to estimate the stability of the aggregated structures is by comparing the number of intra- and inter-molecular HBs in the PTH and  $A\beta$  systems. Both observables have been analyzed extensively for each system in section 5.2.2 for  $A\beta$  and in section 6.2.2 for PTH. Comparisons between the systems considering the temperature dependence of their  $\mathcal{N}_{\text{HB,intra}}(T) = N_{\text{HB,intra}}(T)/N$  and  $\mathcal{N}_{\text{HB,inter}}(T) = N_{\text{HB,inter}}(T)/N$  are shown in fig. 7.1. Given



**Figure 7.1:** Comparison of the number of intra- and inter-molecular hydrogen bonds in the PTH and Aβ systems, relative to the number of beads in the system. Temperature is given in Kelvin, according to the temperature conversions in sections 5.1.1 and 6.1.1.

that Aβ fibrils are more stable than those of PTH, it would be expected that Aβ exhibits a higher number of inter-molecular HBs per residue in the aggregated state. Indeed, 7.1(b) shows that at low temperatures, Aβ(1-42) achieves the highest value of  $\mathcal{N}_{HB,inter}$  at approximately 0.17. However, the values for PTH(1-34), PTH(1-42) and Aβ(1-40) are quite similar, clustering around 0.15. At room temperature ( $T = 300K$ ), the  $\mathcal{N}_{HB,inter}$  of the Aβ systems is slightly below the values of the shorter PTH systems. Only PTH(1-84) shows a significantly lower value of  $\mathcal{N}_{HB,inter} \approx 0.05$ , which remains significantly below that of the Aβ systems, even at  $T = 300K$ .

The comparison between the Aβ systems and PTH(1-84) supports the hypothesis that the reduced stability of the amyloid fibrils of PTH(1-84) is reflected in a lower number of inter-molecular HBs per residues. Furthermore, the trends in  $\mathcal{N}_{HB,inter}$  at  $T = 300K$  suggest that dimers of the shorter PTH(1-34) and PTH(1-42) chains may exhibit stability comparable to or even exceeding that of the Aβ systems. However, it is challenging to establish a direct correlation between the relative number of inter-molecular HBs and the overall stability of the aggregated dimer structure, much less of fully formed fibrils. Nonetheless, the observed difference between the Aβ systems and PTH(1-84) aligns with the *in vivo* behavior of the proteins.

A common feature between the PTH and Aβ dimer systems investigated in this work is their pronounced tendency to form β-sheets. This behavior mirrors the structural characteristics of amyloid fibrils formed by both proteins. Although the dimer structures observed here do not directly resemble the fibril structures, the strong preference for β-sheet formation in the PTH and Aβ dimer systems suggests that dimerization plays a critical early role in the pathway towards amyloid fibril development. The differences in folding and aggregation behavior between the PTH and Aβ systems provide valuable insights into their fibrillation behavior and the stability of their respective amyloid fibrils. This work establishes a foundation for further thermodynamic and structural investigations of their aggregation pathways and structural dynamics.

## Chapter 8

# Outlook

Building upon the findings of this study, a natural extension of the research would be to simulate and analyze larger systems. Investigating the formation of larger oligomers, such as trimers, tetramers, etc., of PolyQ, A $\beta$  and PTH represents a logical next step towards a deeper understanding of fibril structure formation. Additionally, oligomers of amyloidogenic proteins are identified as key contributors to toxicity and resulting cell death in neurodegenerative diseases. This includes PolyQ of HD [224] and A $\beta$  of AD [21]. The PRIME20 model and SAMC method applied in this work could offer valuable insights into the thermodynamics and molecular interactions governing these oligomeric structures.

Conversely, examining smaller systems, particularly protein monomers, could provide complementary insights. For example, the presence of a second chain significantly influenced the secondary structure of monomeric PTH, where the  $\alpha$ -helical structure observed in isolation was almost entirely absent in the dimer systems. To further investigate this phenomenon, future studies should analyze the monomeric structures of PTH and A $\beta$  in greater detail.

Expanding the scope of this research to other amyloidogenic proteins would also be highly beneficial. The versatility of the PRIME20 model makes it well-suited for studying a variety of proteins, offering potential insights into the universal behavior of amyloidogenic systems and fibril formation. Examples of other proteins to explore include the prion protein PrP<sup>Sc</sup> which is associated with Creutzfeldt-Jakob disease [225], and  $\alpha$ -synuclein protein which is associated with Parkinson's disease [226].

In order to simulate larger systems, improvements to both the PRIME20 model and the simulation methodology are essential. As evident from the model description in chapter 2, the initial promise of the PRIME20 model was to be a simple model with only a few parameters. However, the reliance on squeeze parameters to compensate for suboptimal bead sizes makes the model far more complex than initially intended. Simplifications to the model that eliminate the need for these parameters, should be a priority. This would involve a reevaluation of the bead sizes and the interactions between them.

The adaptation of the PRIME20 model for MC simulations also introduces challenges. For example, the criteria for HB formation include auxiliary interactions that are only active once a HB was established. This can potentially lead to ambiguity in potential energy assignment to a configuration when using MC. In certain microstates, the configuration is legal without the HB, while it becomes illegal with an established HB and the corresponding auxiliary interactions turned on. In MD, this apparent ambiguity is solved from the history of the system. However, evolving a system in time for it to have a 'history' is incompatible with the concepts of MC. Revisiting the criteria for HB formation to ensure compatibility with MC methods is therefore critical.

Optimizing the simulation approach itself is equally important. Enhancements that reduce computational cost while improving configurational space sampling are essential to facilitate simulations of larger systems. Parallelization of the simulation code is an important step, alongside

introducing configuration exchange moves between parallel simulation runs in adjacent energy windows to enhance the sampling of the density of states. Additional MC move types could also be implemented to further expand configurational space sampling.

# Bibliography

- [1] European Medicines Agency (EMA). *Legembi*. 2024. URL: <https://www.ema.europa.eu/en/medicines/human/EPAR/leqembi> (visited on 11/25/2024).
- [2] R. Virchow. “Zur Cellulose-Frage”. In: *Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* 6.3 (1854), pp. 416–426.
- [3] M. G. Iadanza, M. P. Jackson, E. W. Hewitt, N. A. Ranson, and S. E. Radford. “A new era for understanding amyloid structures and disease”. In: *Nature Reviews Molecular Cell Biology* 19.12 (2018), pp. 755–773.
- [4] J. D. Sipe and A. S. Cohen. “Review: History of the amyloid fibril”. In: *Journal of Structural Biology* 130.2-3 (2000), pp. 88–98.
- [5] A. Alzheimer. “Über eine eigenartige Erkrankung der Hirnrinde”. In: *Allgemeine Zeitschrift für Psychiatrie und psychisch-gerichtliche Medizin* 64 (1907), pp. 146–147.
- [6] W. T. Astbury and H. J. Woods. “X-Ray studies of the structure of hair, wool, and related fibres. II.-The molecular structure of elastic properties of hair keratin.” In: *Philosophical Transactions of the Royal Society of London* 232.A (1933), pp. 333–394.
- [7] W. T. Astbury, S. Dickinson, and K. Bailey. “The X-ray interpretation of denaturation and the structure of the seed globulins”. In: *Biochemical Journal* 29.10 (1935), pp. 2351–2360.1.
- [8] A. J. Geddes, K. D. Parker, E. D. Atkins, and E. Beighton. ““Cross- $\beta$ ” conformation in proteins”. In: *Journal of Molecular Biology* 32.2 (1968), pp. 343–358.
- [9] E. D. Eanes and G. G. Glenner. “X-Ray Diffraction Studies on Amyloid Filaments”. In: *The Journal of Histochemistry and Cytochemistry* 16.11 (1968), pp. 673–677.
- [10] E. N. Cline, M. A. Bicca, K. L. Viola, and W. L. Klein. “The Amyloid- $\beta$  Oligomer Hypothesis: Beginning of the Third Decade”. In: *Journal of Alzheimer’s Disease* 64.s1 (2018), S567–S610.
- [11] M. Fändrich. “Oligomeric intermediates in amyloid formation: Structure determination and mechanisms of toxicity”. In: *Journal of Molecular Biology* 421.4-5 (2012), pp. 427–440.
- [12] M. H. Gropp, C. L. Klaips, and F. U. Hartl. “Formation of toxic oligomers of polyQ-expanded Huntingtin by prion-mediated cross-seeding”. In: *Molecular Cell* 82.22 (2022), 4290–4306.e11.
- [13] B. Mroczko, M. Groblewska, A. Litman-Zawadzka, J. Kornhuber, and P. Lewczuk. “Amyloid  $\beta$  oligomers (A $\beta$ Os) in Alzheimer’s disease”. In: *Journal of Neural Transmission* 125.2 (2018), pp. 177–191.
- [14] K. L. Viola and W. L. Klein. “Amyloid  $\beta$  oligomers in Alzheimer’s disease pathogenesis, treatment, and diagnosis”. In: *Acta Neuropathologica* 129.2 (2015), pp. 183–206.
- [15] S. Baglioni, F. Casamenti, M. Bucciantini, L. M. Luheshi, N. Taddei, F. Chiti, C. M. Dobson, and M. Stefani. “Prefibrillar amyloid aggregates could be generic toxins in higher organisms”. In: *Journal of Neuroscience* 26.31 (2006), pp. 8160–8167.

- [16] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani. "Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases". In: *Nature* 416.6880 (2002), pp. 507–511.
- [17] C. G. Glabe and R. Kaye. "Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis". In: *Neurology* 66.2 SUPPL. 1 (2006).
- [18] C. G. Glabe. "Common mechanisms of amyloid oligomer pathogenesis in degenerative disease". In: *Neurobiology of Aging* 27.4 (2006), pp. 570–575.
- [19] N. Reixach, S. Deechongkit, X. Jiang, J. W. Kelly, and J. N. Buxbaum. "Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.9 (2004), pp. 2817–2822.
- [20] E. Evangelisti, R. Cascella, M. Becatti, G. Marrazza, C. M. Dobson, F. Chiti, M. Stefani, and C. Cecchi. "Binding affinity of amyloid oligomers to cellular membranes is a generic indicator of cellular dysfunction in protein misfolding diseases". In: *Scientific Reports* 6.April (2016), pp. 1–14.
- [21] M. Serra-Batiste, M. Ninot-Pedrosa, M. Bayoumi, M. Gairí, G. Maglia, and N. Carulla. "A $\beta$ 42 assembles into specific  $\beta$ -barrel pore-forming oligomers in membrane-mimicking environments". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.39 (2016), pp. 10866–10871.
- [22] G. F. Chen, T. H. Xu, Y. Yan, Y. R. Zhou, Y. Jiang, K. Melcher, and H. E. Xu. "Amyloid beta: Structure, biology and structure-based therapeutic development". In: *Acta Pharmacologica Sinica* 38.9 (2017), pp. 1205–1235.
- [23] H. Hampel, J. Hardy, K. Blennow, C. Chen, G. Perry, S. H. Kim, V. L. Villemagne, P. Aisen, M. Vendruscolo, T. Iwatsubo, C. L. Masters, M. Cho, L. Lannfelt, J. L. Cummings, and A. Vergallo. "The Amyloid- $\beta$  Pathway in Alzheimer's Disease". In: *Molecular Psychiatry* 26.10 (2021), pp. 5481–5503.
- [24] J. Wägele, S. De Sio, B. Voigt, J. Balbach, and M. Ott. "How Fluorescent Tags Modify Oligomer Size Distributions of the Alzheimer Peptide". In: *Biophysical Journal* 116.2 (2019), pp. 227–238.
- [25] G. Huntington. "On Chorea". In: *The Medical and Surgical Reporter* 26.789 (1872), p. 317.
- [26] R. H. Walters and R. M. Murphy. "Examining Polyglutamine Peptide Length: A Connection between Collapsed Conformations and Increased Aggregation". In: *Journal of Molecular Biology* 393.4 (2009), pp. 978–992.
- [27] R. Wetzel. "Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence". In: *Journal of Molecular Biology* 421.4-5 (2012), pp. 466–490.
- [28] C. Zuccato, M. Valenza, and E. Cattaneo. "Molecular mechanisms and potential therapeutic targets in Huntington's disease". In: *Physiological Reviews* 90.3 (2010), pp. 905–981.
- [29] M. Huntley and G. B. Golding. "Evolution of simple sequence in proteins". In: *Journal of Molecular Evolution* 51.2 (2000), pp. 131–140.
- [30] A. Kornberg, L. L. Bertsch, J. F. Jackson, and H. G. Khorana. "Enzymatic Synthesis of Deoxyribonucleic Acid, Xvi. Oligonucleotides As". In: *Proceedings of the National Academy of Sciences of the United States of America* 51 (1964), pp. 315–323.
- [31] G. Levinson and G. A. Gutman. "Slipped-strand mispairing: A major mechanism for DNA sequence evolution". In: *Molecular Biology and Evolution* 4.3 (1987), pp. 203–221.

- 
- [32] M. Duyao, A. Lazzarini, A. Falek, W. Koroshetz<sup>8</sup>, D. Sax<sup>2</sup>, E. Bird, J. Vonsattel<sup>1</sup>, E. Bonilla, J. Alvir, J. Bickham Conde', J.-H. Cha, L. Dure, F. Gomez, M. Ramos, J. Sanchez-Ramos, S. Snodgrass, M. De Young, N. Wexler<sup>2</sup>, C. Moscovitz, G. Penchaszadeh, H. Macfarlane, M. Anderson, B. Jenkins, J. Srinidhi, G. Barnes, J. Gusella, and bibinitperiod M. Macdonald. "Trinucleotide repeat length instability and age of onset in Huntington's disease". In: *Nature Genetics* 4.august (1993), pp. 387–392.
- [33] M. E. MacDonald, C. M. Ambrose, M. P. Duyao, R. H. Myers, C. Lin, L. Srinidhi, G. Barnes, S. A. Taylor, M. James, N. Groot, H. MacFarlane, B. Jenkins, M. A. Anderson, N. S. Wexler, J. F. Gusella, G. P. Bates, S. Baxendale, H. Hummerich, S. Kirby, M. North, S. Youngman, R. Mott, G. Zehetner, Z. Sedlacek, A. Poustka, A. M. Frischauf, H. Lehrach, A. J. Buckler, D. Church, L. Doucette-Stamm, M. C. O'Donovan, L. Ribar-Ramirez, M. Shah, V. P. Stanton, S. A. Strobel, K. M. Draths, J. L. Wales, P. Dervan, D. E. Housman, M. Altherr, R. Shiang, L. Thompson, T. Fielder, J. J. Wasmuth, D. Tagle, J. Valdes, L. Elmer, M. Allard, L. Castilla, M. Swaroop, K. Blanchard, F. S. Collins, R. Snell, T. Holloway, K. Gillespie, N. Datson, D. Shaw, and P. S. Harper. "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes". In: *Cell* 72.6 (1993), pp. 971–983.
- [34] R. G. Snell, J. C. Macmillan, J. P. Cheadle, I. Fenton, L. P. Lazarou, P. Davies, M. E. Macdonald, J. F. Gusella, P. S. Harper, and D. J. Shaw. "Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease". In: *Nature Genetics* 4.4 (1993), pp. 393–397.
- [35] K. Kiebertz, M. MacDonald, C. Shih, A. Feigin, K. Steinberg, K. Bordwell, C. Zimmerman, J. Srinidhi, J. Sotack, J. Gusella, and I. Shoulson. "Trinucleotide repeat length and progression of illness in Huntington's disease". In: *Journal of Medical Genetics* 31.11 (1994), pp. 872–874.
- [36] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel. "Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent". In: *Nature Structural and Molecular Biology* 18.3 (2011), pp. 328–336.
- [37] K. Kar, C. L. Hoop, K. W. Drombosky, M. A. Baker, R. Kodali, I. Arduini, P. C. van der Wel, W. S. Horne, and R. Wetzel. " $\beta$ -Hairpin-Mediated Nucleation of Polyglutamine Amyloid Formation". In: *Journal of Molecular Biology* 425.7 (2013), pp. 1183–1197.
- [38] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, and R. V. Pappu. "Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.45 (2006), pp. 16764–16769.
- [39] D. Otzen and R. Riek. "Functional amyloids". In: *Cold Spring Harbor Perspectives in Biology* 11.12 (2019).
- [40] R. R. MacGregor and D. V. Cohn. "The intracellular pathway for parathormone biosynthesis and secretion". In: *Clinical Orthopaedics and Related Research* 137 (1978), pp. 244–258.
- [41] J. F. Habener and J. T. Potts. "Biosynthesis of parathyroid hormone". In: *The New England Journal Of Medicine* 299 (1978), pp. 580–585.
- [42] K. M. Wiren, J. T. Potts, and H. M. Kronenberg. "Importance of the propeptide sequence of human preproparathyroid hormone for signal sequence function". In: *Journal of Biological Chemistry* 263.36 (1988), pp. 19771–19777.
- [43] Z. S. Agus, L. B. Gardner, L. H. Beck, and M. Goldberg. "Effects of parathyroid hormone on renal tubular calcium and phosphate handling". In: *American Journal of Physiology* 1224.5 (1973), pp. 1143–1148.
-

- [44] S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. R. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale, and R. Riek. “Functional amyloids as natural storage of peptide hormones in pituitary secretory granules”. In: *Science* 325.5938 (2009), pp. 328–332.
- [45] F. Liang, C. L. Liu, and R. J. Carroll. “Stochastic approximation in Monte Carlo computation”. In: *Journal of the American Statistical Association* 102.477 (Mar. 2007), pp. 305–320.
- [46] M. Cheon, I. Chang, and C. K. Hall. “Extending the PRIME model for protein aggregation to all 20 amino acids”. In: *Proteins: Structure, Function and Bioinformatics* 78.14 (2010), pp. 2950–2960.
- [47] R. D. Hills and C. L. Brooks. “Insights from coarse-grained go models for protein folding and dynamics”. In: *International Journal of Molecular Sciences* 10.3 (2009), pp. 889–905.
- [48] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski. “Coarse-Grained Protein Models and Their Applications”. In: *Chemical Reviews* 116.14 (2016), pp. 7898–7936.
- [49] N. Rathore and J. J. De Pablo. “Monte Carlo simulation of proteins through a random walk in energy space”. In: *Journal of Chemical Physics* 116.16 (2002), pp. 7225–7230.
- [50] M. P. Taylor, W. Paul, and K. Binder. “All-or-none proteinlike folding transition of a flexible homopolymer chain”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 79.5 (2009), pp. 1–4.
- [51] V. Tozzini. “Coarse-grained models for proteins”. In: *Current Opinion in Structural Biology* 15.2 (2005), pp. 144–150.
- [52] V. Tozzini. *Minimalist models for proteins: A comparative analysis*. Vol. 43. 3. 2010, pp. 333–371.
- [53] A. Böker and W. Paul. “Thermodynamics and Conformations of Single Polyalanine, Polyserine, and Polyglutamine Chains within the PRIME20 Model”. In: *The Journal of Physical Chemistry B* 126.38 (Sept. 2022), pp. 7286–7297.
- [54] M. Cheon, I. Chang, and C. K. Hall. “Spontaneous formation of twisted A $\beta$  16-22 Fibrils in Large-Scale Molecular-Dynamics Simulations”. In: *Biophysical Journal* 101.10 (2011), pp. 2493–2501.
- [55] M. Cheon, I. Chang, and C. K. Hall. “Influence of temperature on formation of perfect tau fragment fibrils using PRIME20/DMD simulations”. In: *Protein Science* 21.10 (2012), pp. 1514–1527.
- [56] C. Lauer and W. Paul. “Dimerization of Polyglutamine within the PRIME20 Model using Stochastic Approximation Monte Carlo”. In: *Macromolecular Theory and Simulations* 2200075 (2023), pp. 1–11.
- [57] Y. Wang, Q. Shao, and C. K. Hall. “N-terminal prion protein peptides (PrP(120-144)) form parallel in-register  $\beta$ -sheets via multiple nucleation-dependent pathways”. In: *Journal of Biological Chemistry* 291.42 (2016), pp. 22093–22105.
- [58] Y. Wang and C. K. Hall. “Seeding and cross-seeding fibrillation of N-terminal prion protein peptides PrP(120–144)”. In: *Protein Science* 27.7 (2018), pp. 1304–1313.
- [59] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithm to Applications*. 2nd ed. Academic Press, 2002.
- [60] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids: Second edition*. 2nd ed. Oxford University Press, 2017.
- [61] A. Böker. “Simulations of protein thermodynamics and structures with the PRIME20 model”. PhD thesis. Martin-Luther-Universität Halle-Wittenberg, 2019.

- 
- [62] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. 4th ed. New York, NY, USA: Freeman, W. H., 2005.
- [63] J. J. Berzelius. *Bref utgifna af Kungl: svenska vetenskapsakademien genom HG Söderbaum: Lettres publiées au nom de l'Académie royale des sciences de Suède*. Almqvist & Wiksells, 1916.
- [64] H. B. Vickery. "The origin of the word protein". In: *The Yale journal of biology and medicine* 22.5 (1950), pp. 387–393.
- [65] L. Pauling and R. Corey. "Helical Configurations of Polypeptide Chains N-". In: *Proceedings of the National Academy of Sciences* 37.5 (1951), pp. 235–240.
- [66] L. Pauling and R. B. Corey. "The pleated sheet, a new layer configuration of polypeptide chains." In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (1951), pp. 251–256.
- [67] L. Pauling, R. B. Corey, and H. R. Branson. "The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain." In: *Proceedings of the National Academy of Sciences of the United States of America* 37.4 (1951), pp. 205–211.
- [68] E. N. Baker and R. E. Hubbard. "Hydrogen bonding in globular proteins". In: *Progress in Biophysics and Molecular Biology* 44.2 (1984), pp. 97–179.
- [69] K. Gunasekaran, C. Ramakrishnan, and P. Balaram. " $\beta$ -Hairpins in proteins revisited: Lessons for de novo design". In: *Protein Engineering* 10.10 (1997), pp. 1131–1141.
- [70] E. J. Milner-White and R. Poet. "Four classes of beta-hairpins in proteins." In: *The Biochemical journal* 240.1 (1986), pp. 289–292.
- [71] B. Sibanda and J. Thornton. " $\beta$ -Hairpin families in globular proteins". In: *Nature* 316 (1985), pp. 170–174.
- [72] L. Bragg, J. C. Kendrew, and M. F. Perutz. "Polypeptide chain configurations in crystalline proteins". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 203 (1950), pp. 321–357.
- [73] B. K. Ho, A. Thomas, and R. Brasseur. "Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the  $\alpha$ -helix". In: *Protein Science* 12.11 (2003), pp. 2508–2522.
- [74] S. Hovmöller, T. Zhou, and T. Ohlson. "Conformations of amino acids in proteins research papers". In: *Biological Crystallography* D58 (2002), pp. 768–776.
- [75] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. "Stereochemistry of polypeptide chain configurations". In: *Journal of Molecular Biology* 7.1 (1963), pp. 95–99.
- [76] G. N. Ramachandran and V. Sasisekharan. "Conformation of Polypeptides and Proteins". In: *Advances in Protein Chemistry* 23.C (1968), pp. 283–437.
- [77] R. Schneider, M. C. Schumacher, H. Mueller, D. Nand, V. Klaukien, H. Heise, D. Riedel, G. Wolf, E. Behrmann, S. Raunser, R. Seidel, M. Engelhard, and M. Baldus. "Structural characterization of polyglutamine fibrils by solid-state NMR spectroscopy". In: *Journal of Molecular Biology* 412.1 (2011), pp. 121–136.
- [78] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. "A structural model for Alzheimer's  $\beta$ -amyloid fibrils based on experimental constraints from solid state NMR". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.26 (2002), pp. 16742–16747.
-

- [79] M. Gopalswamy, A. Kumar, J. Adler, M. Baumann, M. Henze, S. T. Kumar, M. Fändrich, H. A. Scheidt, D. Huster, and J. Balbach. “Structural characterization of amyloid fibrils from the human parathyroid hormone”. In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1854.4 (2015), pp. 249–257.
- [80] T. R. Jahn, O. S. Makin, K. L. Morris, K. E. Marshall, P. Tian, P. Sikorski, and L. C. Serpell. “The Common Architecture of Cross- $\beta$  Amyloid”. In: *Journal of Molecular Biology* 395.4 (2010), pp. 717–727.
- [81] I. Gerroff, A. Milchev, K. Binder, and W. Paul. “A new off-lattice Monte Carlo model for polymers: A comparison of static and dynamic properties with the bond-fluctuation model and application to random media”. In: *The Journal of Chemical Physics* 98.8 (1993), pp. 6526–6539.
- [82] A. Milchev and K. Binder. “Static and Dynamic Properties of Adsorbed Chains at Surfaces: Monte Carlo Simulation of a Bead-Spring Model”. In: *Macromolecules* 29.1 (1996), pp. 343–354.
- [83] C. Vanderzande. *Lattice models of polymers*. 11. Cambridge University Press, 1998.
- [84] K. A. Dill. “Theory for the Folding and Stability of Globular Proteins”. In: *Biochemistry* 24.6 (1985), pp. 1501–1509.
- [85] T. E. Gartner and A. Jayaraman. “Modeling and Simulations of Polymers: A Roadmap”. In: *Macromolecules* 52.3 (2019), pp. 755–786.
- [86] G. Khaira, M. Doxastakis, A. Bowen, J. Ren, H. S. Suh, T. Segal-Peretz, X. Chen, C. Zhou, A. F. Hannon, N. J. Ferrier, V. Vishwanath, D. F. Sunday, R. Gronheid, R. J. Kline, J. J. De Pablo, and P. F. Nealey. “Derivation of Multiple Covarying Material and Process Parameters Using Physics-Based Modeling of X-ray Data”. In: *Macromolecules* 50.19 (2017), pp. 7783–7793.
- [87] M. Levitt and A. Washel. “Computational simulation of protein folding”. In: *Nature* 253 (1975), pp. 694–698.
- [88] M. W. Matsen. “The standard Gaussian model for block copolymer melts”. In: *Journal of Physics Condensed Matter* 14.2 (2002).
- [89] H. Abe and N. Gō. “Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins”. In: *Biopolymers* 20.5 (1981), pp. 1013–1031.
- [90] H. Taketomi, Y. Ueda, and N. Gō. “Studies on protein folding. unfolding and fluctuations by computer simulation: I. The effect of specific amino acid sequence represented by specific inter-unit interactions”. In: *International Journal of Peptide and Protein Research* 7.6 (1975), pp. 445–459.
- [91] J. Karanicolas and C. L. Brooks. “Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions”. In: *Journal of Molecular Biology* 334.2 (2003), pp. 309–325.
- [92] J. I. Sułkowska and M. Cieplak. “Selection of optimal variants of Gō-like models of proteins through studies of stretching”. In: *Biophysical Journal* 95.7 (2008), pp. 3174–3191.
- [93] P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. “Biomolecular dynamics: Order-disorder transitions and energy landscapes”. In: *Reports on Progress in Physics* 75.7 (2012).
- [94] B. Werlich. “Intramolekulare Strukturbildung durch Steifigkeitsvariation und Auswirkung der Anwesenheit von spezifischen Wechselwirkungen”. PhD thesis. Martin-Luther-Universität Halle-Wittenberg, 2018.

- 
- [95] S. Sun. “Reduced representation model of protein structure prediction: Statistical potential and genetic algorithms”. In: *Protein Science* 2.5 (1993), pp. 762–785.
- [96] S. Takada, Z. Luthey-Schulten, and P. G. Wolynes. “Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer”. In: *Journal of Chemical Physics* 110.23 (1999), pp. 11616–11629.
- [97] A. Voegler Smith and C. K. Hall. “ $\alpha$ -helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model”. In: *Proteins: Structure, Function and Genetics* 44.3 (2001), pp. 344–360.
- [98] A. V. Smith and C. K. Hall. “Assembly of a tetrameric  $\alpha$ -helical bundle: Computer simulations on an intermediate-resolution protein model”. In: *Proteins: Structure, Function and Genetics* 44.3 (2001), pp. 376–391.
- [99] A. V. Smith and C. K. Hall. “Protein refolding versus aggregation: Computer simulations on an intermediate-resolution protein model”. In: *Journal of Molecular Biology* 312.1 (2001), pp. 187–202.
- [100] A. Voegler Smith. “Simulations of Protein Refolding and Aggregation Using a Novel Intermediate-Resolution Protein Model”. PhD thesis. North Carolina State University, 2001.
- [101] H. D. Nguyen, A. J. Marchut, and C. K. Hall. “Solvent effects on the conformational transition of a model polyalanine peptide”. In: *Protein Science* 13.11 (2004), pp. 2909–2924.
- [102] H. A. Lorentz. “Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase”. In: *Annalen der Physik* 248.1 (1881), pp. 127–136.
- [103] C. B. Anfinsen. “Principles that Govern the Folding of Protein Chains”. In: *Science* 181.4096 (1973).
- [104] D. T. Jones, W. R. Taylor, and J. M. Thornton. “A New Approach to Protein Folding Calculations”. In: 358.July (1992), pp. 86–89.
- [105] W. Krauth and M. Mezard. “Learning algorithms with optimal stability in neural networks”. In: *Journal of Physics A: General Physics* 20.11 (1987).
- [106] S. Miyazawa and R. L. Jernigan. “Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation”. In: *Macromolecules* 18.3 (1985), pp. 534–552.
- [107] S. Miyazawa and R. L. Jernigan. “Residue – Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading”. In: *Journal of Molecular Biology* 256.3 (Mar. 1996), pp. 623–644.
- [108] C. K. Hall and V. A. Wagoner. “Computational Approaches to Fibril Structure and Formation”. In: *Methods in Enzymology* 412.1997 (2006), pp. 338–365.
- [109] A. J. Marchut and C. K. Hall. “Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations”. In: *Biophysical Journal* 90.12 (2006), pp. 4574–4584.
- [110] A. J. Marchut and C. K. Hall. “Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides”. In: *Computational Biology and Chemistry* 30.3 (2006), pp. 215–218.
- [111] A. J. Marchut and C. K. Hall. “Effects of Chain Length on the Aggregation of Model Polyglutamine Peptides: Molecular Dynamics Simulations”. In: *Proteins: Structure, Function and Bioinformatics* 66 (2007), pp. 96–109.
-

- [112] M. Cheon, C. K. Hall, and I. Chang. “Structural Conversion of A $\beta$ 17–42 Peptides from Disordered Oligomers to U-Shape Protofilaments via Multiple Kinetic Pathways”. In: *PLoS Computational Biology* 11.5 (2015), pp. 1–23.
- [113] M. Cheon, M. Kang, and I. Chang. “Polymorphism of fibrillar structures depending on the size of assembled A $\beta$ 17-42 peptides”. In: *Scientific Reports* 6.November (2016), pp. 1–11.
- [114] D. C. Latshaw, M. Cheon, and C. K. Hall. “Effects of macromolecular crowding on amyloid beta (16-22) aggregation using coarse-grained simulations”. In: *Journal of Physical Chemistry B* 118.47 (2014), pp. 13513–13526.
- [115] D. C. Latshaw and C. K. Hall. “Effects of Hydrophobic Macromolecular Crowders on Amyloid  $\beta$  (16-22) Aggregation”. In: *Biophysical Journal* 109.1 (2015), pp. 124–134.
- [116] Q. Shao and C. K. Hall. “A Discontinuous Potential Model for Protein–Protein Interactions”. In: *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015* (2016), pp. 1–20.
- [117] V. A. Wagoner, M. Cheon, I. Chang, and C. K. Hall. “Computer simulation study of amyloid fibril formation by palindromic sequences in prion peptides”. In: *Proteins: Structure, Function and Bioinformatics* 79.7 (2011), pp. 2132–2145.
- [118] V. A. Wagoner, M. Cheon, I. Chang, and C. K. Hall. “Fibrillization propensity for short designed hexapeptides predicted by computer simulation”. In: *Journal of Molecular Biology* 416.4 (2012), pp. 598–609.
- [119] V. A. Wagoner, M. Cheon, I. Chang, and C. K. Hall. “Impact of sequence on the molecular assembly of short amyloid peptides”. In: *Proteins: Structure, Function and Bioinformatics* 82.7 (2014), pp. 1469–1483.
- [120] L. Landau and E. Lifshitz. *Statistical Physics*. Third Revi. Oxford: Pergamon Press Ltd., 1980.
- [121] W. Nolting. *Grundkurs Theoretische Physik 6 - Statistische Physik*. 1998.
- [122] J. Bernoulli. *Ostwald’s Klassiker der exakten Wissenschaften. Nr.108. Wahrscheinlichkeitsrechnung (Ars conjectandi) (1713)*. Ed. by R. Haussner. Leipzig: Wilhelm Engelmann, 1899.
- [123] J. W. Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. New York, USA: C. Scribner’s sons, 1902.
- [124] R. Eckhardt. “Stan Ulam, John von Neumann and the Monte Carlo Method”. In: *Los Alamos Science* 15 (1987), pp. 131–137.
- [125] S. Ulam and N. Metropolis. “The Monte Carlo Method”. In: *Journal of the American Statistical Association* 44.247 (1949), pp. 335–341.
- [126] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.
- [127] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. January. 2010.
- [128] J. Torrie, G. M Valleau. “Nonphysical Smpling Distributions on Monte Carlo Free-Energy Estimation: Umbrella Sampling”. In: *J. Comput. Phys.* 23 (1977), pp. 187–199.
- [129] U. H. Hansmann and Y. Okamoto. “Finite-size scaling of helix–coil transitions in polyalanine studied by multicanonical simulations”. In: *The Journal of Chemical Physics* 110.2 (1999), pp. 1267–1276.

- 
- [130] W. Janke. “Computer Simulation Studies of Polymer Adsorption and Aggregation - From Flexible to Stiff”. In: *Physics Procedia* 68 (2015), pp. 69–79.
  - [131] W. Janke and W. Paul. “Thermodynamics and structure of macromolecules from flat-histogram Monte Carlo simulations”. In: *Soft Matter* 12.3 (2016), pp. 642–657.
  - [132] D. P. Landau and F. Wang. “Determining the density of states for classical statistical models by a flat-histogram random walk”. In: *Computer Physics Communications* 147.1 (2002), pp. 674–677.
  - [133] F. Wang and D. P. Landau. “Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram”. In: *Phys. Rev. E* 64.5 (2001), p. 56101.
  - [134] F. Wang and D. P. Landau. “Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States”. In: *Phys. Rev. Lett.* 86.10 (2001), pp. 2050–2053.
  - [135] B. Werlich, M. P. Taylor, and W. Paul. “Wang-Landau and stochastic approximation Monte Carlo for semi-flexible polymer chains”. In: *Physics Procedia* 57.C (2014), pp. 82–86.
  - [136] B. Werlich, T. Shakirov, M. P. Taylor, and W. Paul. “Stochastic approximation Monte Carlo and Wang-Landau Monte Carlo applied to a continuum polymer model”. In: *Computer Physics Communications* 186 (2015), pp. 65–70.
  - [137] E. Filimonova, V. Ivanov, and T. Shakirov. “Distinguishing noisy crystal symmetries in coarse-grained computer simulations: New procedures for noise reduction and lattice reconstruction”. In: *ArXiv Preprint* (2025). URL: <https://arxiv.org/abs/2404.15539>.
  - [138] F. Liang. “A theory on flat histogram Monte Carlo algorithms”. In: *Journal of Statistical Physics* 122.3 (2006), pp. 511–529.
  - [139] F. Liang. “An Overview of Stochastic Approximation Monte Carlo”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.4 (2014), pp. 240–254.
  - [140] T. Shakirov. “Flat-histogram algorithms: optimal parameters and extended application”. In: *ArXiv Preprint* (2024), pp. 1–14. URL: <http://arxiv.org/abs/2402.05653>.
  - [141] D. Maltseva, S. Zablotskiy, J. Martemyanova, V. Ivanov, T. Shakirov, and W. Paul. “Diagrams of states of single flexible-semiflexible multi-block copolymer chains: A flat-histogram Monte Carlo study”. In: *Polymers* 11.5 (2019), pp. 1–20.
  - [142] T. Shakirov, S. Zablotskiy, A. Böker, V. Ivanov, and W. Paul. “Comparison of Boltzmann and Gibbs entropies for the analysis of single-chain phase transitions”. In: *European Physical Journal: Special Topics* 226.4 (2017), pp. 705–723.
  - [143] M. J. Williams. “Microcanonical Analysis of Helical Homopolymers: Exploring the Density of States and Structural Characteristics”. In: *Polymers* 15.19 (2023).
  - [144] B. Werlich, T. Shakirov, M. P. Taylor, and W. Paul. “Stochastic approximation Monte Carlo and Wang-Landau Monte Carlo applied to a continuum polymer model”. In: *Computer Physics Communications* 186 (2015), pp. 65–70.
  - [145] T. Shakirov. “Convergence estimation of flat-histogram algorithms based on simulation results”. In: *Computer Physics Communications* 228 (2018), pp. 38–43.
  - [146] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. New York, NY, USA, 1988.
  - [147] L. Verlet. “Computer ”Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: *Phys. Rev.* 159.1 (1967), pp. 98–103.
  - [148] E. Filimonova. “Polymer crystallization at different substrates : flat histogram Monte Carlo simulation”. PhD thesis. Martin-Luther-Universität Halle-Wittenberg, 2024.
-

- [149] Python Software Foundation. *Python Language Reference*. URL: <http://www.python.org>.
- [150] ISO. *ISO/IEC 14882:2017 Information technology — Programming languages — C++*. Fifth. Dec. 2017, p. 1605.
- [151] Free Software Foundation. *Bash*. 2007. URL: <https://www.gnu.org/software/bash/>.
- [152] P. M. Spencer Kimball and the GIMP Development Team. *GNU Image Manipulation Program*. Version 2.10.36. URL: <https://www.gimp.org/>.
- [153] Inkscape Project. *Inkscape*. Version 1.3.2. 2020. URL: <https://inkscape.org>.
- [154] Schrödinger LLC. *The PyMOL Molecular Graphics System*. Version 2.5.0. 2015. URL: <https://pymol.org>.
- [155] A. D. Sokal. “Monte Carlo Methods for the Self-Avoiding Walk”. In: *Monte Carlo and Molecular Dynamics Simulations in Polymer Science1*. Oxford University Press, 1995.
- [156] D. Frenkel and P. B. Warren. “Gibbs, Boltzmann, and negative temperatures”. In: *American Journal of Physics* 83.2 (2015), pp. 163–170.
- [157] S. Hilbert, P. Hänggi, and J. Dunkel. “Thermodynamic laws in isolated systems”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 90.6 (2014), pp. 1–23.
- [158] D. H. E. Gross. *Microcanonical Thermodynamics - Phase Transitions in "Small" Systems*. Singapore: World Scientific Publishing Co. Ptc. Ltd., 2001.
- [159] K. Qi and M. Bachmann. “Classification of Phase Transitions by Microcanonical Inflection-Point Analysis”. In: *Physical Review Letters* 120.18 (2018), pp. 1–5.
- [160] S. Schnabel, D. T. Seaton, D. P. Landau, and M. Bachmann. “Microcanonical entropy inflection points: Key to systematic understanding of transitions in finite systems”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 84.1 (2011), pp. 1–4.
- [161] G. Jaeger. “The ehrenfest classification of phase transitions: Introduction and evolution”. In: *Archive for History of Exact Sciences* 53.1 (1998), pp. 51–81.
- [162] J. G. Brankov. *Leuven Notes in Mathematical and Theoretical Physics - Volume 8 - Series A: Mathematical physics - Introduction to finite-size scaling*. Leuven: Leuven University Press, 1996.
- [163] V. A. Ivanov, W. Paul, and K. Binder. “Finite chain length effects on the coil-globule transition of stiff-chain macromolecules: A Monte Carlo simulation”. In: *Journal of Chemical Physics* 109.13 (1998), pp. 5659–5669.
- [164] M. Chen, M. Tsai, W. Zheng, and P. G. Wolynes. “The Aggregation Free Energy Landscapes of Polyglutamine Repeats”. In: *Journal of the American Chemical Society* 138.46 (2016), pp. 15197–15203.
- [165] M. Han and U. H. Hansmann. “Replica exchange molecular dynamics of the thermodynamics of fibril growth of Alzheimers A42 peptide”. In: *Journal of Chemical Physics* 135.6 (2011).
- [166] T. Takeda and D. K. Klimov. “Replica exchange simulations of the thermodynamics of A $\beta$  fibril growth”. In: *Biophysical Journal* 96.2 (2009), pp. 442–452.
- [167] B. Zhang-Haagen, R. Biehl, L. Nagel-Steger, A. Radulescu, D. Richter, and D. Willbold. “Monomeric Amyloid beta peptide in hexafluoroisopropanol detected by small angle neutron scattering”. In: *PLoS ONE* 11.2 (2016), pp. 1–12.

- 
- [168] S. Nag, B. Sarkar, A. Bandyopadhyay, B. Sahoo, V. K. Sreenivasan, M. Kombrabail, C. Muralidharan, and S. Maiti. “Nature of the amyloid- $\beta$  monomer and the monomer-oligomer equilibrium”. In: *Journal of Biological Chemistry* 286.16 (2011), pp. 13827–13833.
- [169] P. Anand, F. S. Nandel, and U. H. Hansmann. “The Alzheimer  $\beta$ -amyloid (A $\beta$ 1-39) dimer in an implicit solvent”. In: *Journal of Chemical Physics* 129.19 (2008).
- [170] W. M. Berhanu and U. H. Hansmann. “Structure and dynamics of amyloid- $\beta$  segmental polymorphisms”. In: *PLoS ONE* 7.7 (2012).
- [171] I. Bertini, L. Gonnelli, C. Luchinat, J. Mao, and A. Nesi. “A new structural model of A $\beta$  40 fibrils”. In: *Journal of the American Chemical Society* 133.40 (2011), pp. 16013–16022.
- [172] M. Garvey, M. Baumann, M. Wulff, S. T. Kumar, D. Markx, I. Morgado, U. Knüpfer, U. Horn, C. Mawrin, M. Fändrich, and J. Balbach. “Molecular architecture of A $\beta$  fibrils grown in cerebrospinal fluid solution and in a cell culture model of A $\beta$  plaque formation”. In: *Amyloid* 23.2 (2016), pp. 76–85.
- [173] Y. R. Huang and R. T. Liu. “The toxicity and polymorphism of  $\beta$ -amyloid oligomers”. In: *International Journal of Molecular Sciences* 21.12 (2020), pp. 1–19.
- [174] A. R. Lam, D. B. Teplow, H. E. Stanley, and B. Urbanc. “Effects of the arctic (E22→G) mutation on amyloid  $\beta$ -protein folding: Discrete molecular dynamics study”. In: *Journal of the American Chemical Society* 130.51 (2008), pp. 17413–17422.
- [175] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek. “3D structure of Alzheimer’s amyloid- $\beta$ (1-42) fibrils”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.48 (2005), pp. 17342–17347.
- [176] Y. Miller, B. Ma, and R. Nussinov. “Polymorphism in alzheimer A $\beta$  amyloid organization reflects conformational selection in a rugged energy landscape”. In: *Chemical Reviews* 110.8 (2010), pp. 4820–4838.
- [177] C. Sachse, M. Fändrich, and N. Grigorieff. “Paired  $\beta$ -sheet structure of an A $\beta$ (1-40) amyloid fibril revealed by electron microscopy”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.21 (2008), pp. 7462–7466.
- [178] M. Schmidt, C. Sachse, W. Richter, C. Xu, M. Fändrich, and N. Grigorieff. “Comparison of Alzheimer A $\beta$ (1-40) and A $\beta$ (1-42) amyloid fibrils reveals similar protofilament structures”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.47 (2009), pp. 19813–19818.
- [179] N. G. Sgourakis, Y. Yan, S. A. McCallum, C. Wang, and A. E. Garcia. “The Alzheimer’s Peptides A $\beta$ 40 and 42 Adopt Distinct Conformations in Water: A Combined MD / NMR Study”. In: *Journal of Molecular Biology* 368.5 (2007), pp. 1448–1457.
- [180] M. Thunecke, A. Lobbia, U. Kosciessa, T. Dyrks, A. E. Oakley, J. Turner, W. Saenger, and Y. Georgalis. “Aggregation of A $\beta$  Alzheimer’s disease-related peptide studied by dynamic light scattering”. In: *Journal of Peptide Research* 52.6 (1998), pp. 509–517.
- [181] B. Urbanc, M. Betnel, L. Cruz, G. Bltan, and D. B. Teplow. “Elucidation of amyloid  $\beta$ -protein oligomerization mechanisms: Discrete molecular dynamics study”. In: *Journal of the American Chemical Society* 132.12 (2010), pp. 4266–4280.
- [182] J. P. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. B. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. R. Sawaya, and D. Eisenberg. “Molecular basis for amyloid- $\beta$  polymorphism”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.41 (2011), pp. 16938–16943.
-

- [183] S. Côté, P. Derreumaux, and N. Mousseau. “Distinct morphologies for amyloid beta protein monomer: A $\beta$ 1-40, A $\beta$ 1-42, and A $\beta$ 1-40(D23N)”. In: *Journal of Chemical Theory and Computation* 7.8 (2011), pp. 2584–2592.
- [184] S. Côté, R. Laghaei, P. Derreumaux, and N. Mousseau. “Distinct Dimerization for Various Alloforms of the Amyloid-Beta Protein: A $\beta$ 1-40, A $\beta$ 1-42, and A $\beta$ 1-40(D23N)”. In: *Journal of Physical Chemistry B* 116.13 (2012), pp. 4043–4055.
- [185] J. J. Balbach, A. T. Petkova, N. A. Oyler, O. N. Antzutkin, D. J. Gordon, S. C. Meredith, and R. Tycko. “Supramolecular structure in full-length Alzheimer’s  $\beta$ -amyloid fibrils: Evidence for a parallel  $\beta$ -sheet organization from solid-state nuclear magnetic resonance”. In: *Biophysical Journal* 83.2 (2002), pp. 1205–1216.
- [186] M. A. Wälti, F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall, A. Böckmann, P. Güntert, B. H. Meier, and R. Riek. “Atomic-resolution structure of a disease-relevant A $\beta$ (1-42) amyloid fibril”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.34 (2016), E4976–E4984.
- [187] K. A. Ball, A. H. Phillips, D. E. Wemmer, and T. Head-Gordon. “Differences in  $\beta$ -strand populations of monomeric A $\beta$ 40 and A $\beta$ 42”. In: *Biophysical Journal* 104.12 (2013), pp. 2714–2724.
- [188] L. Hou, H. Shao, Y. Zhang, H. Li, N. K. Menon, E. B. Neuhaus, J. M. Brewer, I. J. L. Byeon, D. G. Ray, M. P. Vitek, T. Iwashita, R. A. Makula, A. B. Przybyla, and M. G. Zagorski. “Solution NMR Studies of the A $\beta$ (1-40) and A $\beta$ (1-42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation”. In: *Journal of the American Chemical Society* 126.7 (2004), pp. 1992–2005.
- [189] R. Riek, P. Güntert, H. Döbeli, B. Wipf, and K. Wüthrich. “NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, A $\beta$ (1-40)ox and A $\beta$ (1-42)ox”. In: *European Journal of Biochemistry* 268.22 (2001), pp. 5930–5936.
- [190] S. Zhang, K. Iwata, M. J. Lachenmann, J. W. Peng, S. Li, E. R. Stimson, Y. A. Lu, A. M. Felix, J. E. Maggio, and J. P. Lee. “The Alzheimer’s peptide A $\beta$  adopts a collapsed coil structure in water”. In: *Journal of Structural Biology* 130.2-3 (2000), pp. 130–141.
- [191] P. Anand, F. S. Nandel, and U. H. Hansmann. “The Alzheimer’s  $\beta$  amyloid (A $\beta$ 1-39) monomer in an implicit solvent”. In: *Journal of Chemical Physics* 128.16 (2008).
- [192] V. H. Man, P. H. Nguyen, and P. Derreumaux. “High-Resolution Structures of the Amyloid- $\beta$  1-42 Dimers from the Comparison of Four Atomistic Force Fields”. In: *Journal of Physical Chemistry B* 121.24 (2017), pp. 5977–5987.
- [193] S. Mitternacht, I. Staneva, T. Hard, and A. Irback. “Comparing the folding free-energy landscapes of A $\beta$ 42 variants with different aggregation properties”. In: *Proteins: Structure, Function and Bioinformatics* 78.12 (2010), pp. 2600–2608.
- [194] S. Mitternacht, I. Staneva, T. Härd, and A. Irback. “Monte Carlo study of the formation and conformational properties of dimers of A $\beta$ 42 variants”. In: *Journal of Molecular Biology* 410.2 (2011), pp. 357–367.
- [195] M. Cheon, I. Chang, and C. K. Hall. “Spontaneous formation of twisted A $\beta$  16-22 Fibrils in Large-Scale Molecular-Dynamics Simulations”. In: *Biophysical Journal* 101.10 (2011), pp. 2493–2501.
- [196] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Madsen, C. Riek, R. Grothe, and D. Eisenberg. “Structure of the cross- $\beta$  spine of amyloid-like fibrils”. In: *Nature* 435.7043 (2005), pp. 773–778.
- [197] A. T. Petkova, W. M. Yau, and R. Tycko. “Experimental constraints on quaternary structure in Alzheimer’s  $\beta$ -amyloid fibrils”. In: *Biochemistry* 45.2 (2006), pp. 498–512.

- 
- [198] S. N. Malkov, M. V. Živković, M. V. Beljanski, M. B. Hall, and S. D. Zarić. “A reexamination of the propensities of amino acids towards a particular secondary structure: Classification of amino acids based on their chemical structure”. In: *Journal of Molecular Modeling* 14.8 (2008), pp. 769–775.
- [199] K. Imai and S. Mitaku. “Mechanisms of secondary structure breakers in soluble proteins”. In: *Biophysics* 1 (2005), pp. 55–65.
- [200] P. C. Lyu, M. I. Liff, L. A. Marky, and N. R. Kallenbach. “Side Chain Contributions to the Stability of Alpha-Helical Structure in Peptides”. In: *Science* 250.4981 (1990), pp. 669–673.
- [201] C. A. Kim and J. M. Berg. “Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide”. In: *Nature* 362 (1993), pp. 267–270.
- [202] J. S. Merkel and L. Regan. “Aromatic rescue of glycine in  $\beta$  sheets”. In: *Folding and Design* 3.6 (1998), pp. 449–456.
- [203] D. L. Minor and P. S. Kim. “Measurement of the  $\beta$ -sheet-forming propensities of amino acids”. In: *Nature* 367.6464 (1994), pp. 660–663.
- [204] K. T. O’Neil and W. F. DeGrado. “A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids”. In: *Science* 250.4981 (1990), pp. 646–651.
- [205] C. K. Smith, L. Regan, and J. M. Withka. “A Thermodynamic Scale for the  $\beta$ -Sheet Forming Tendencies of the Amino Acids”. In: *Biochemistry* 33.18 (1994), pp. 5510–5517.
- [206] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax. “TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts”. In: *Journal of Biomolecular NMR* 44.4 (2009), pp. 213–223.
- [207] D. Frishman and P. Argos. “Knowledge-based protein secondary structure assignment”. In: *Proteins: Structure, Function, and Bioinformatics* 23.4 (1995), pp. 566–579.
- [208] O. Gursky and S. Aleshkov. “Temperature-dependent  $\beta$ -sheet formation in  $\beta$ -amyloid A $\beta$ 1-40 peptide in water: Uncoupling  $\beta$ -structure folding from aggregation”. In: *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* 1476.1 (2000), pp. 93–102.
- [209] J. Jarvet, P. Damberg, J. Danielsson, I. Johansson, L. E. Eriksson, and A. Gräslund. “A left-handed 3<sub>1</sub> helical conformation in the Alzheimer A $\beta$ (12-28) peptide”. In: *FEBS Letters* 555.2 (2003), pp. 371–374.
- [210] S. Vivekanandan, J. R. Brender, S. Y. Lee, and A. Ramamoorthy. “A partially folded structure of amyloid-beta(1-40) in an aqueous environment”. In: *Biochemical and Biophysical Research Communications* 411.2 (2011), pp. 312–316.
- [211] H. B. Brewer, T. Fairwell, R. Ronan, G. W. Sizemore, and C. D. Arnaud. “Human parathyroid hormone: amino-acid sequence of the amino-terminal residues 1-34.” In: *Proceedings of the National Academy of Sciences of the United States of America* 69.12 (1972), pp. 3585–3588.
- [212] R. Brommage, C. E. Hotchkiss, C. J. Lees, M. W. Stancill, J. M. Hock, and C. P. Jerome. “Daily treatment with human recombinant parathyroid hormone-(1-34), LY333334, for 1 year increases bone mass in ovariectomized monkeys”. In: *Journal of Clinical Endocrinology and Metabolism* 84.10 (1999), pp. 3757–3763.
- [213] G. W. Tregear, J. V. Rietschoten, E. Greene, H. T. Keutmann, H. D. Niall, B. Reit, J. A. Parsons, and J. T. Potts. “Bovine parathyroid hormone: Minimum chain length of synthetic peptide required for biological activity”. In: *Endocrinology* 93.6 (1973), pp. 1349–1353.
-

- [214] U. H. Hansmann. “Generalized-ensemble simulations of the human parathyroid hormone fragment PTH(1-34)”. In: *Journal of Chemical Physics* 120.1 (2004), pp. 417–422.
- [215] S. Y. Kim and W. Kwak. “All-atom simulation study of protein PTH(1-34) by using the Wang-Landau sampling method”. In: *Journal of the Korean Physical Society* 65.11 (2014), pp. 1733–1737.
- [216] C. Y. Liu, X. Li, W. Y. Chen, L. C. Chang, Y. F. Chen, H. L. Chen, Y. S. Sun, H. Y. Lai, and E. W. Huang. “PEGylation site-dependent structural heterogeneity study of MonoPEGylated human parathyroid hormone fragment hPTH(1-34)”. In: *Langmuir* 30.38 (2014), pp. 11421–11427.
- [217] J. A. Barden and B. E. Kemp. “NMR Solution Structure of Human Parathyroid Hormone(1–34)”. In: *Biochemistry* 32.28 (1993), pp. 7126–7132.
- [218] U. C. Marx, S. Austermann, P. Bayer, K. Adermann, A. Ejchart, H. Sticht, S. Walter, F. X. Schmid, R. Jaenicke, W. G. Forssmann, and P. Rösch. “Structure of human parathyroid hormone 1-37 in solution”. In: *Journal of Biological Chemistry* 270.25 (1995), pp. 15194–15202.
- [219] K. J. Willis and A. G. Szabo. “Conformation of Parathyroid Hormone: Time-Resolved Fluorescence Studies”. In: *Biochemistry* 31.37 (1992), pp. 8924–8931.
- [220] A. Paschold, M. Schäffler, X. Miao, L. Gardon, S. Krüger, H. Heise, M. I. Röhr, M. Ott, B. Strodel, and W. H. Binder. “Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides”. In: *Bioconjugate Chemistry* 35.7 (2024), pp. 981–995.
- [221] U. C. Marx, K. Adermann, P. Bayer, W. G. Forssmann, and P. Rösch. “Solution structures of human parathyroid hormone fragments hPTH(1-34) and hPTH(1-39) and bovine parathyroid hormone fragment bPTH(1-37)”. In: *Biochemical and Biophysical Research Communications* 267.1 (2000), pp. 213–220.
- [222] O. Oren, R. Taube, and N. Papo. “Amyloid  $\beta$  structural polymorphism, associated toxicity and therapeutic strategies”. In: *Cellular and Molecular Life Sciences* 78.23 (2021), pp. 7185–7198.
- [223] R. Tycko. “Amyloid Polymorphism: Structural Basis and Neurobiological Relevance”. In: *Neuron* 86.3 (2015), pp. 632–645.
- [224] T. Takahashi, S. Katada, and O. Onodera. “Polyglutamine diseases: Where does toxicity come from? What is toxicity? Where are we going?” In: *Journal of Molecular Cell Biology* 2.4 (2010), pp. 180–191.
- [225] B. Sikorska, R. Knight, J. W. Ironside, and P. Liberski. “Creutzfeldt-Jakob Disease”. In: *Neurodegenerative Diseases*. Ed. by S. I. Ahmad. New York, NY: Springer US, 2012, pp. 76–90.
- [226] L. Stefanis. “ $\alpha$ -Synuclein in Parkinson’s disease”. In: *Cold Spring Harbor perspectives in medicine* 2.2 (2012), a009399.

# Appendix A

## Parameters of PRIME20

**Table A.1:** Side chain bond and pseudo-bond lengths (in Å) and bead masses (relative to the mass of CH<sub>3</sub>).

	C $\alpha$ -R	NH-R	CO-R	Mass
L	2.625	3.290	3.500	3.799
I	2.400	3.050	3.300	3.799
V	2.002	2.775	2.959	2.866
M	3.400	3.800	4.050	4.998
F	3.425	3.650	4.050	6.061
Y	3.843	4.050	4.300	7.126
W	3.881	4.100	4.350	8.660
H	3.160	3.450	3.830	5.394
S	1.967	2.650	2.800	2.064
T	1.981	2.650	2.900	2.997
N	2.510	3.050	3.350	3.862
Q	3.300	3.750	4.000	4.795
D	2.500	3.100	3.250	3.860
E	3.180	3.780	3.930	4.793
K	3.550	4.050	4.250	4.865
R	4.200	4.500	4.800	6.728
C	2.350	2.800	3.100	3.130
P	1.926	1.851	2.995	2.800
A	1.600	2.500	2.560	1.000

**Table A.2:** Side chain square-well depth (energies). The grouping is in accordance with the Prime20 classification scheme

	LIV	M	F	Y	W	H	ST	NQ	DE	KR	C	P	A
LIV	-0.200	-0.200	-0.203	-0.203	-0.203	0.015	0.015	0.015	0.015	0.015	-0.139	0.015	-0.148
M		-0.200	-0.203	-0.210	-0.210	-0.116	-0.116	-0.116	0.015	-0.116	-0.139	0.015	-0.148
F			-0.205	-0.205	-0.205	0.015	0.015	0.015	0.015	0.015	-0.139	0.015	-0.148
Y				-0.201	-0.201	-0.086	-0.086	-0.086	-0.086	-0.086	-0.116	0.015	-0.148
W					-0.205	-0.086	-0.116	-0.086	-0.086	0.015	-0.116	0.015	-0.148
H						-0.080	-0.086	-0.080	-0.086	-0.086	-0.116	0.074	0.074
ST							-0.086	-0.086	-0.086	-0.086	-0.116	0.074	0.074
NQ								-0.080	-0.086	-0.086	-0.116	0.074	0.074
DE									-0.086	-0.086	-0.116	0.074	0.074
KR									0.253	-0.136	-0.116	0.074	0.074
C										0.073	-0.116	0.074	0.074
P											-0.585	0.015	-0.139
A												0.074	0.074
													-0.084

**Table A.3:** Side chain hard-sphere bead diameter in Å.

	L	I	V	M	F	Y	W	H	S	T	N	Q	D	E	K	R	C	P	A
L	3.4	3.4	3.0	3.6	3.4	3.2	3.4	3.2	3.0	3.2	3.4	3.5	3.0	3.3	3.5	3.4	3.4	3.5	2.7
I		3.3	3.3	3.6	3.4	3.0	3.2	3.1	2.6	3.0	2.8	3.1	3.4	3.2	2.9	3.6	3.3	3.5	2.9
V			3.3	3.0	3.2	3.0	2.9	3.1	2.8	2.8	3.1	3.3	3.0	3.1	3.1	3.1	2.9	3.3	2.7
M				3.7	3.2	3.2	3.2	3.6	3.2	3.6	3.5	3.4	3.6	3.3	3.7	3.7	3.4	3.7	2.9
F					3.3	3.2	3.4	2.9	2.9	2.8	2.7	3.3	3.1	3.3	3.5	3.3	3.2	3.1	2.4
Y						3.0	3.2	3.1	2.9	3.2	3.3	3.4	2.8	3.3	3.5	3.1	2.9	3.3	2.7
W							3.7	3.2	2.7	3.3	2.8	3.4	3.2	3.5	3.5	3.1	2.9	3.3	2.7
H								3.4	2.6	2.9	3.4	3.3	2.8	3.3	3.4	3.5	2.8	3.7	3.1
S									2.5	2.9	3.0	2.7	2.8	2.9	3.0	3.0	2.8	3.2	2.3
T										2.9	3.1	3.3	3.1	3.1	3.1	3.2	2.7	2.6	2.6
N											3.3	3.5	3.2	3.1	3.2	2.9	3.1	3.3	2.8
Q												3.6	2.8	2.9	3.4	3.6	3.1	3.6	3.0
D													3.4	2.9	3.0	3.0	3.2	3.2	2.6
E														3.2	3.4	3.1	2.7	3.5	2.9
K															3.5	3.9	2.7	3.6	3.3
R																3.2	3.3	3.0	3.0
C																	2.1	3.0	2.8
P																		3.1	2.9
A																			2.7

**Table A.4:** Side chain square-well diameter in Å.

	L	I	V	M	F	Y	W	H	S	T	N	Q	D	E	K	R	C	P	A
L	6.4	6.5	6.2	6.5	6.6	6.7	6.9	6.5	6.3	6.2	6.4	6.3	6.5	6.4	6.5	6.8	6.1	6.3	5.6
I		6.6	6.4	6.7	6.6	6.8	6.8	6.6	6.4	6.4	6.6	6.6	6.5	6.6	6.7	6.7	6.4	6.4	5.7
V			6.3	6.4	6.5	6.5	6.6	6.2	6.2	6.4	6.3	6.5	6.3	6.5	6.6	6.8	6.0	6.3	6.1
M				6.7	6.5	6.6	7.0	6.5	6.4	6.4	6.4	6.4	6.7	6.4	6.4	6.6	6.3	6.2	5.8
F					6.8	6.8	7.0	6.5	6.2	6.6	6.5	6.6	6.7	6.8	6.9	6.9	6.4	6.5	5.9
Y						7.0	7.0	6.9	6.5	6.4	6.7	6.7	6.9	6.8	6.7	7.0	6.5	6.4	5.7
W							7.4	7.1	6.3	6.5	6.9	6.7	6.9	6.9	6.5	6.9	6.4	6.3	5.5
H								6.7	6.3	6.3	6.5	6.6	6.6	6.4	6.6	6.9	6.2	6.3	5.5
S									6.4	6.0	6.2	6.0	6.1	6.0	6.1	6.3	6.3	6.1	5.9
T										6.5	6.3	6.4	6.2	6.4	6.5	6.8	6.1	6.6	6.2
N											6.3	6.4	6.5	6.4	6.5	6.6	6.2	6.2	5.6
Q												6.6	6.3	6.6	6.7	6.9	6.1	6.5	5.8
D													6.5	6.6	6.3	6.5	6.2	6.3	5.6
E														6.7	6.4	6.6	6.1	6.4	5.9
K															6.9	6.8	6.4	6.7	6.0
R																7.2	6.3	6.8	6.1
C																	6.2	6.0	5.9
P																		6.5	6.2
A																			5.4

---

**Table A.5:** Changed diameters for interactions between side chains and backbone (after applying squeeze parameters *sqz6-10*) for all amino acids.

	<i>sqz6</i>	<i>sqz7</i>	<i>sqz8</i>	<i>sqz9</i>	<i>sqz10</i>
	$R_i\text{-CO}_{i-1}$	$R_i\text{-NH}_{i+1}$	$R_i\text{-C}_{\alpha i+1}$	$R_i\text{-C}_{\alpha i-1}$	$R_i\text{-C}_{\alpha i-2}$
L	3.918	3.724	4.863	4.936	5.001
I	3.740	3.626	4.867	4.867	4.994
V	3.570	3.378	4.635	4.754	5.002
M	4.205	4.032	5.067	5.206	5.017
F	3.991	3.973	4.827	4.780	5.040
Y	4.208	4.246	4.978	4.898	5.042
W	4.460	4.187	4.963	5.180	4.986
H	3.886	3.838	4.790	4.766	4.945
S	3.331	3.128	4.380	4.507	4.944
T	3.447	3.290	4.573	4.617	5.007
N	3.607	3.565	4.680	4.633	4.791
Q	4.139	3.996	5.062	5.134	5.000
D	3.751	3.435	4.558	4.785	4.860
E	4.175	3.997	5.074	5.162	4.996
K	4.384	4.191	5.163	5.323	4.974
R	4.827	4.651	5.535	5.703	4.978
C	3.516	3.350	4.501	4.560	4.913
P	3.133	3.298	4.665	3.884	4.773
A	3.312	3.000	4.353	4.598	4.997

---

## Appendix B

### Additional Figures for the Polyglutamine Systems

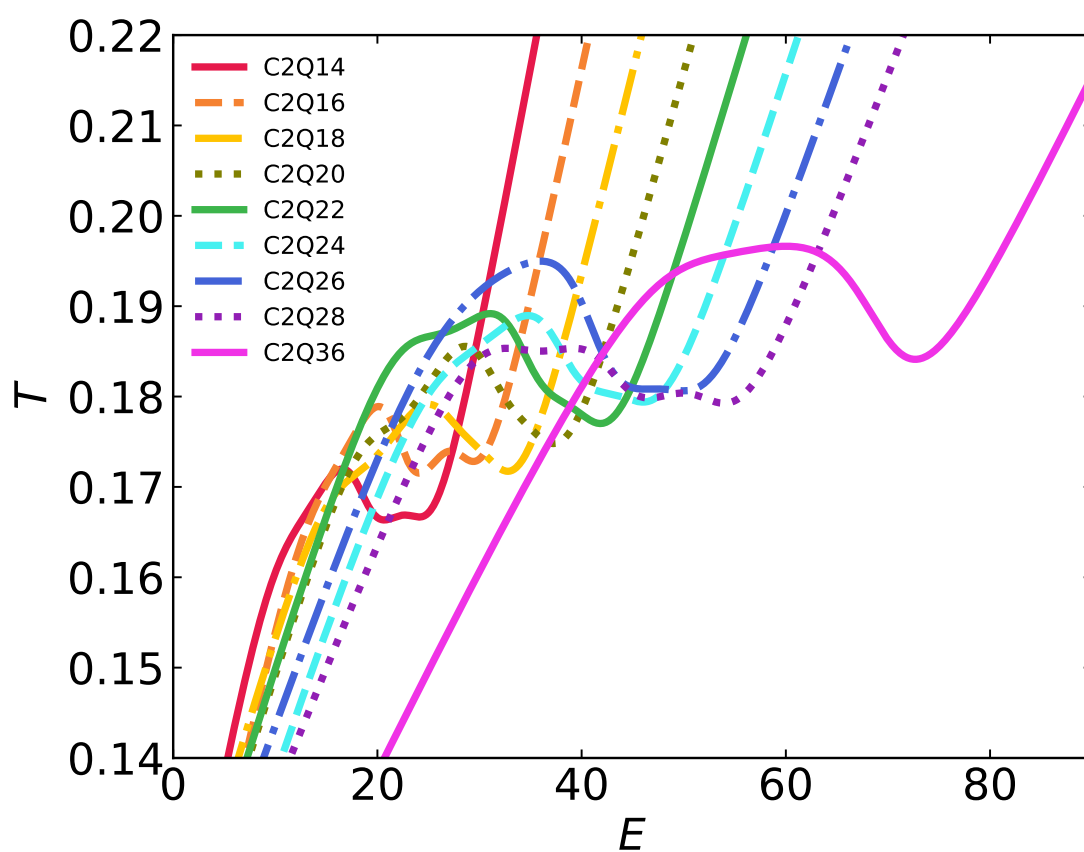


Figure B.1: PolyQ temperatures

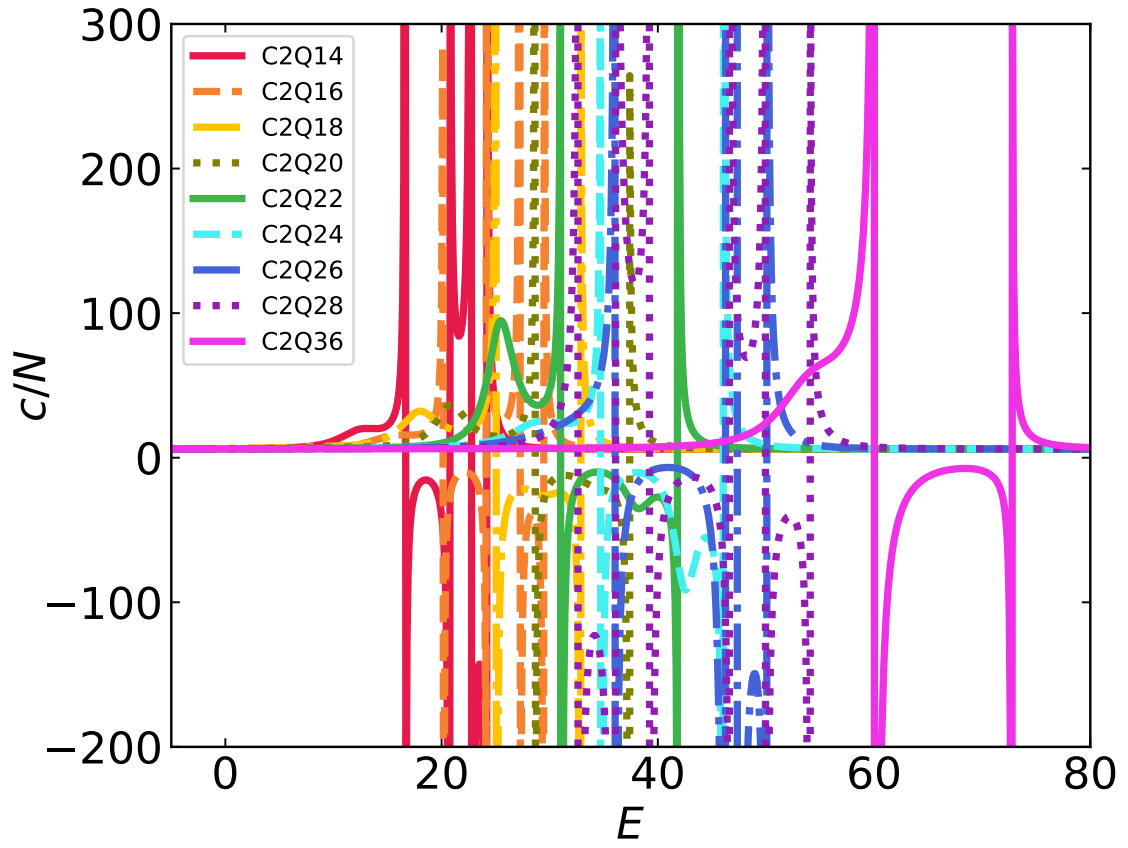


Figure B.2: PolyQ heat capacities

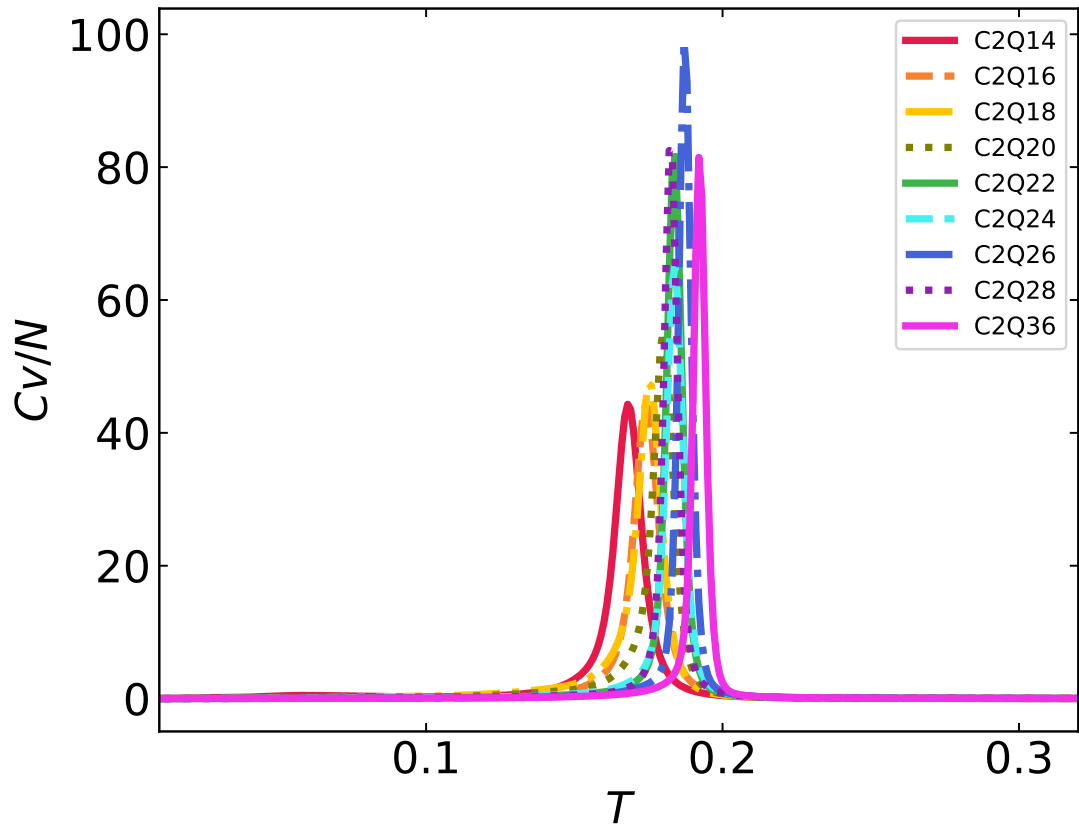


Figure B.3: PolyQ canonical heat capacities

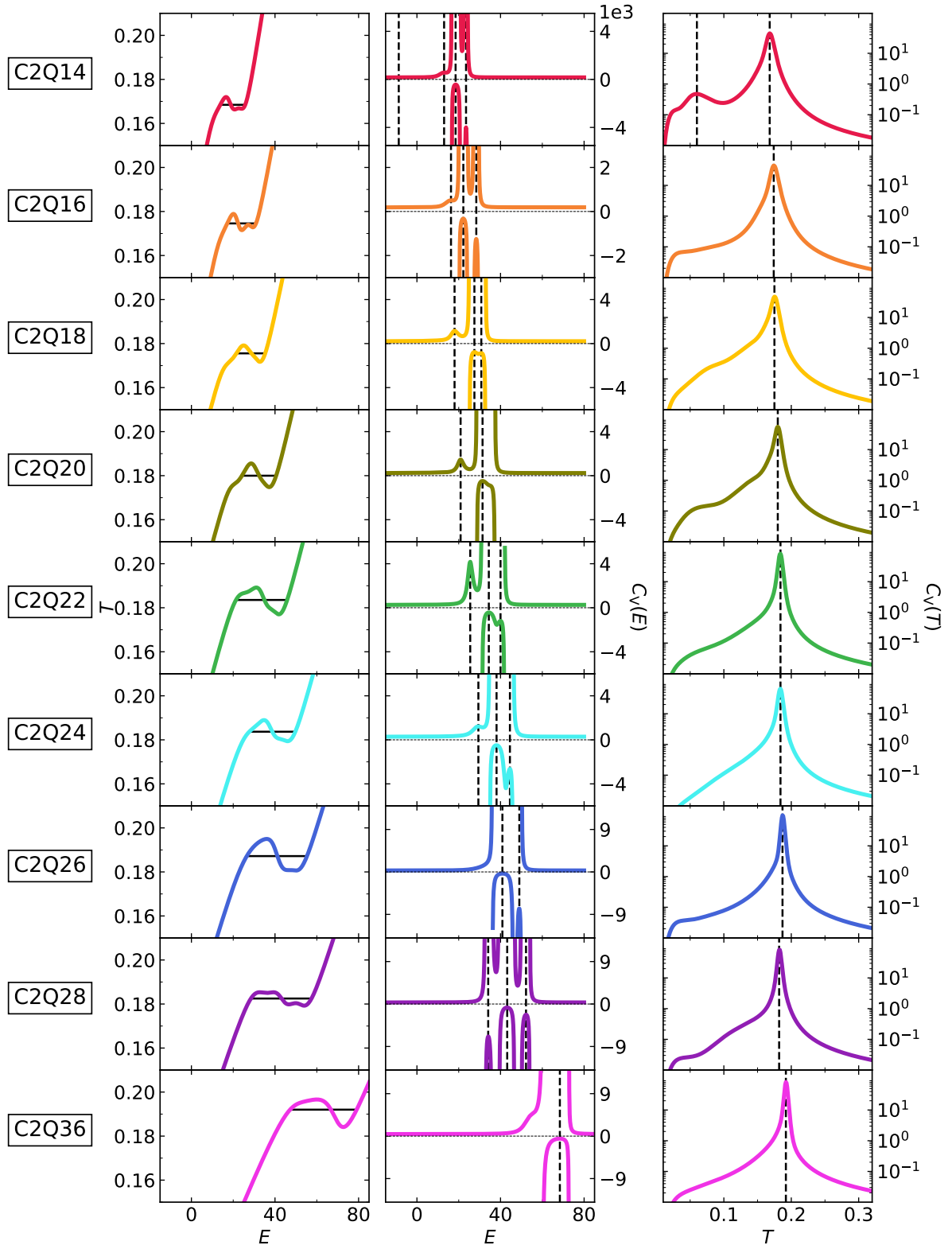
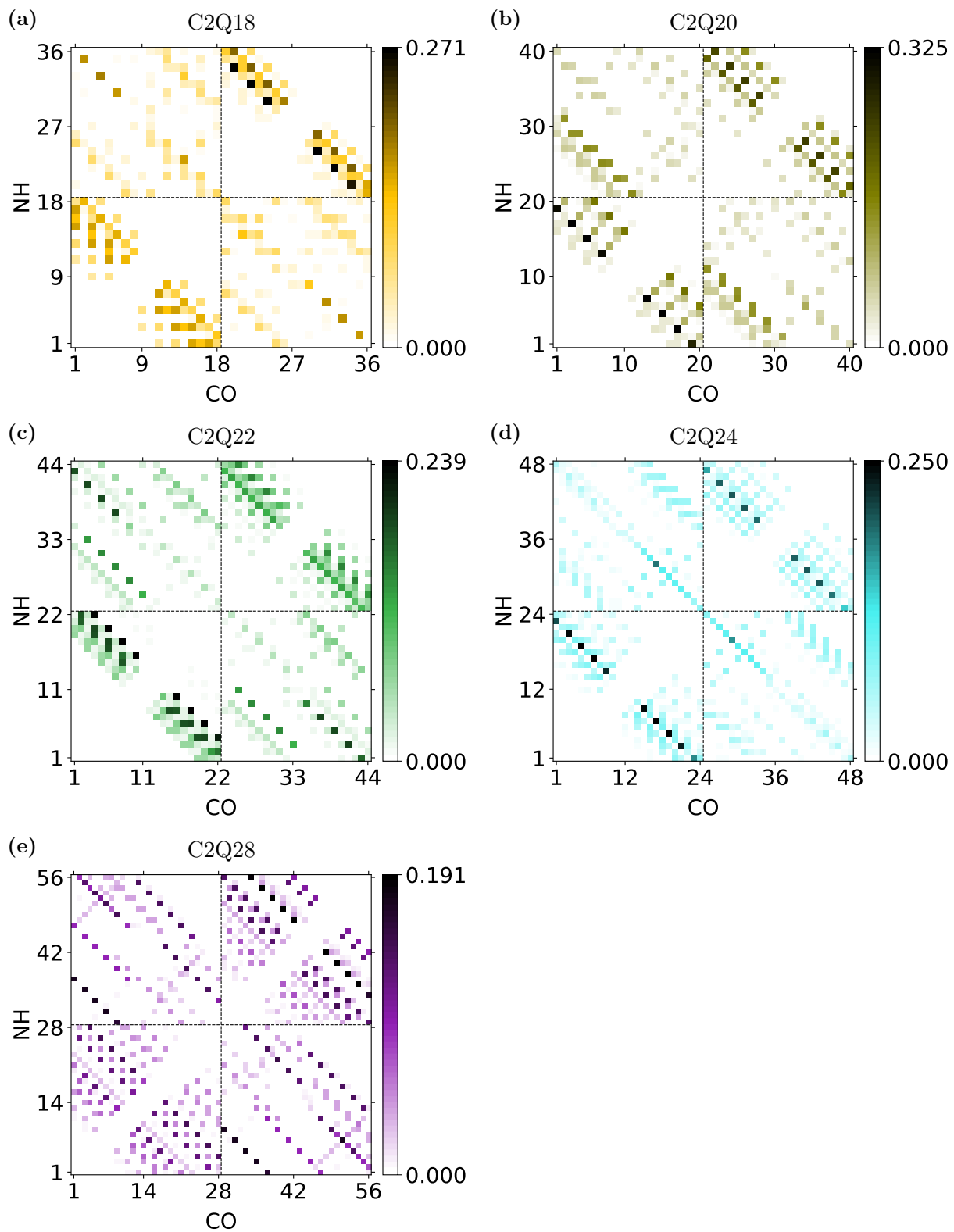


Figure B.4: PolyQ thermodynamic master plot.



**Figure B.5:** Intra-molecular hydrogen bond contact probability maps for C2Q18, C2Q20, C2Q22, C2Q24 and C2Q28, at  $T = 0.140$ .

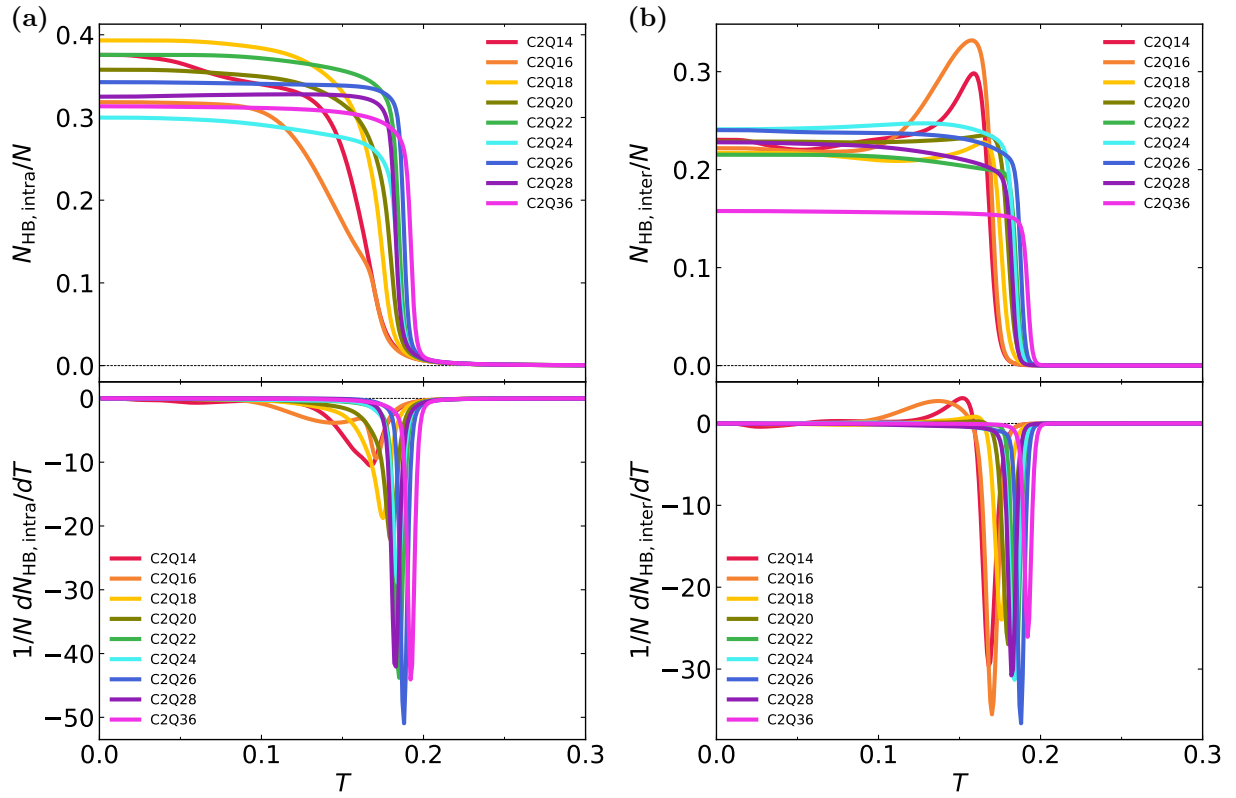


Figure B.6: Intra- and inter-molecular HBs

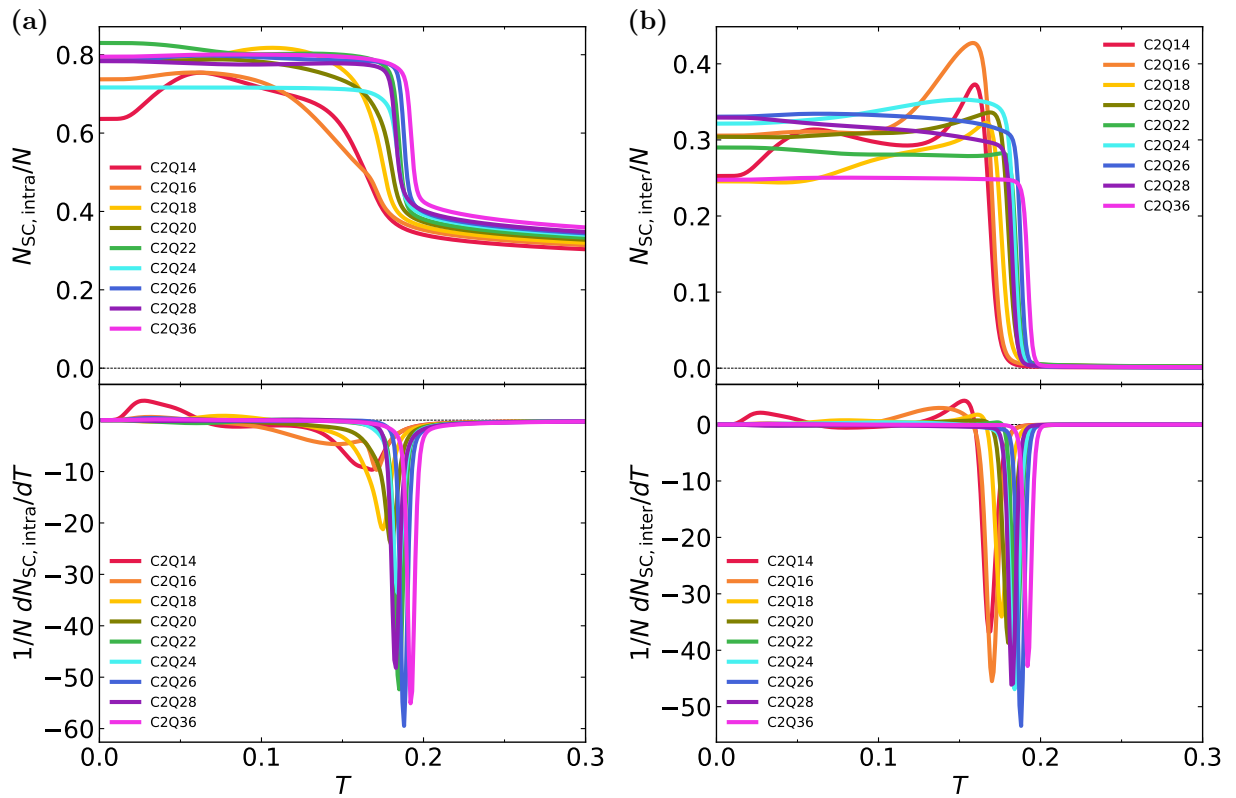
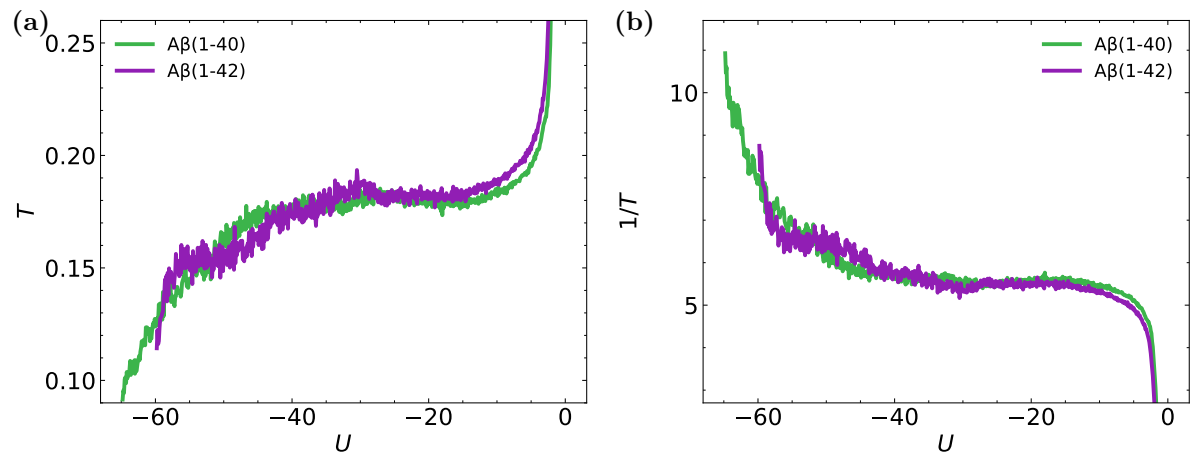


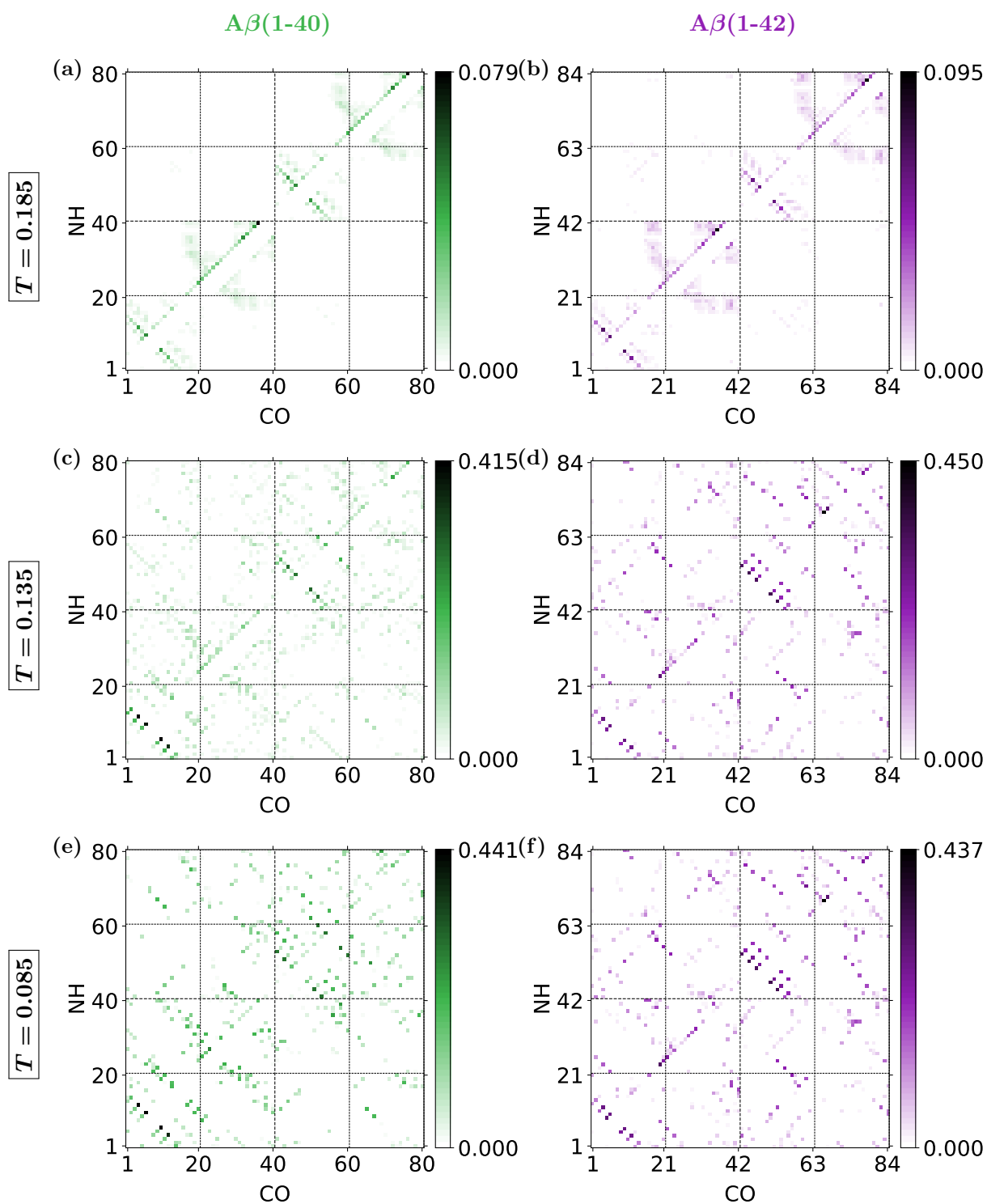
Figure B.7: Intra- and inter-molecular side chain contacts

## Appendix C

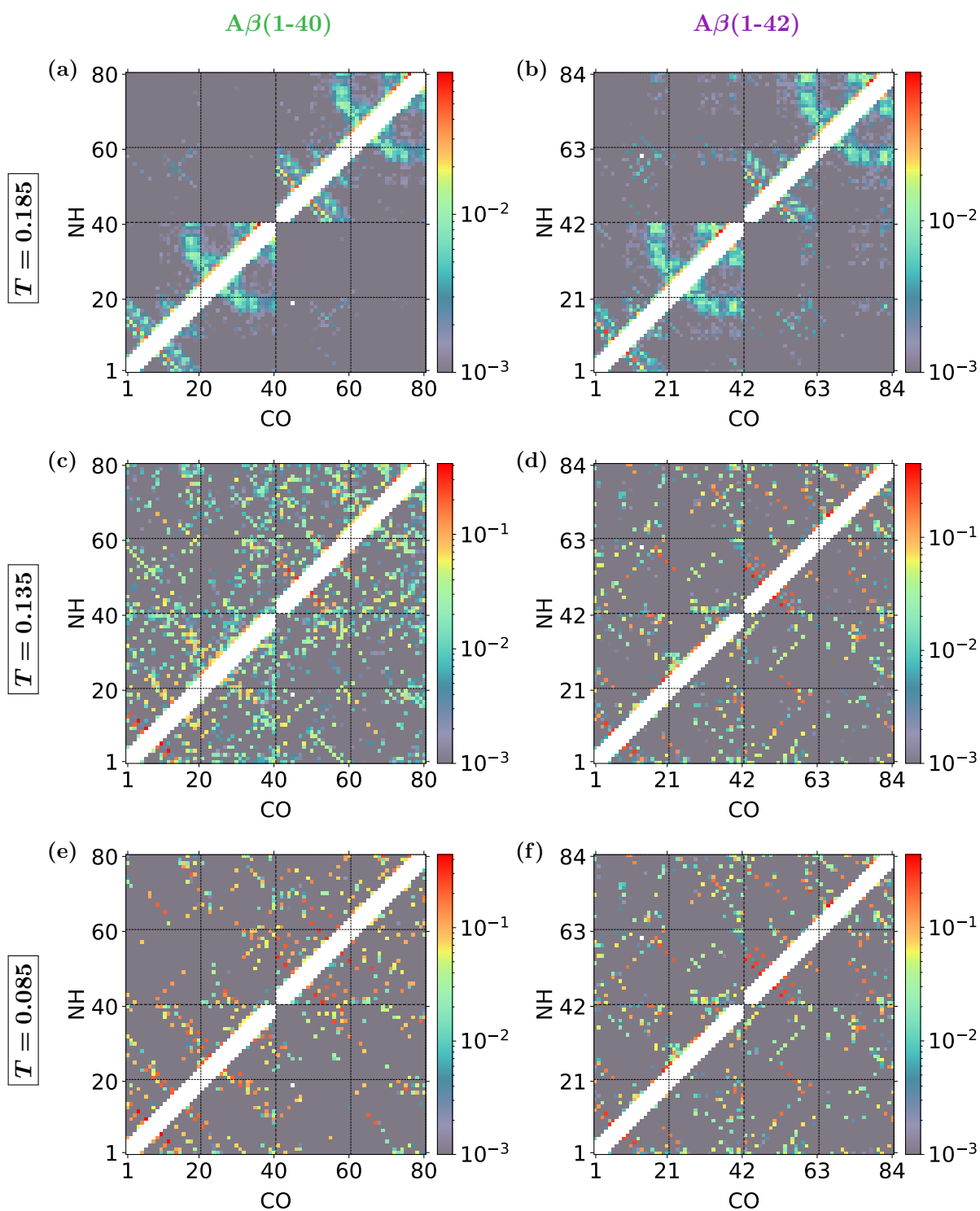
# Additional Figures for the Amyloid Beta Systems



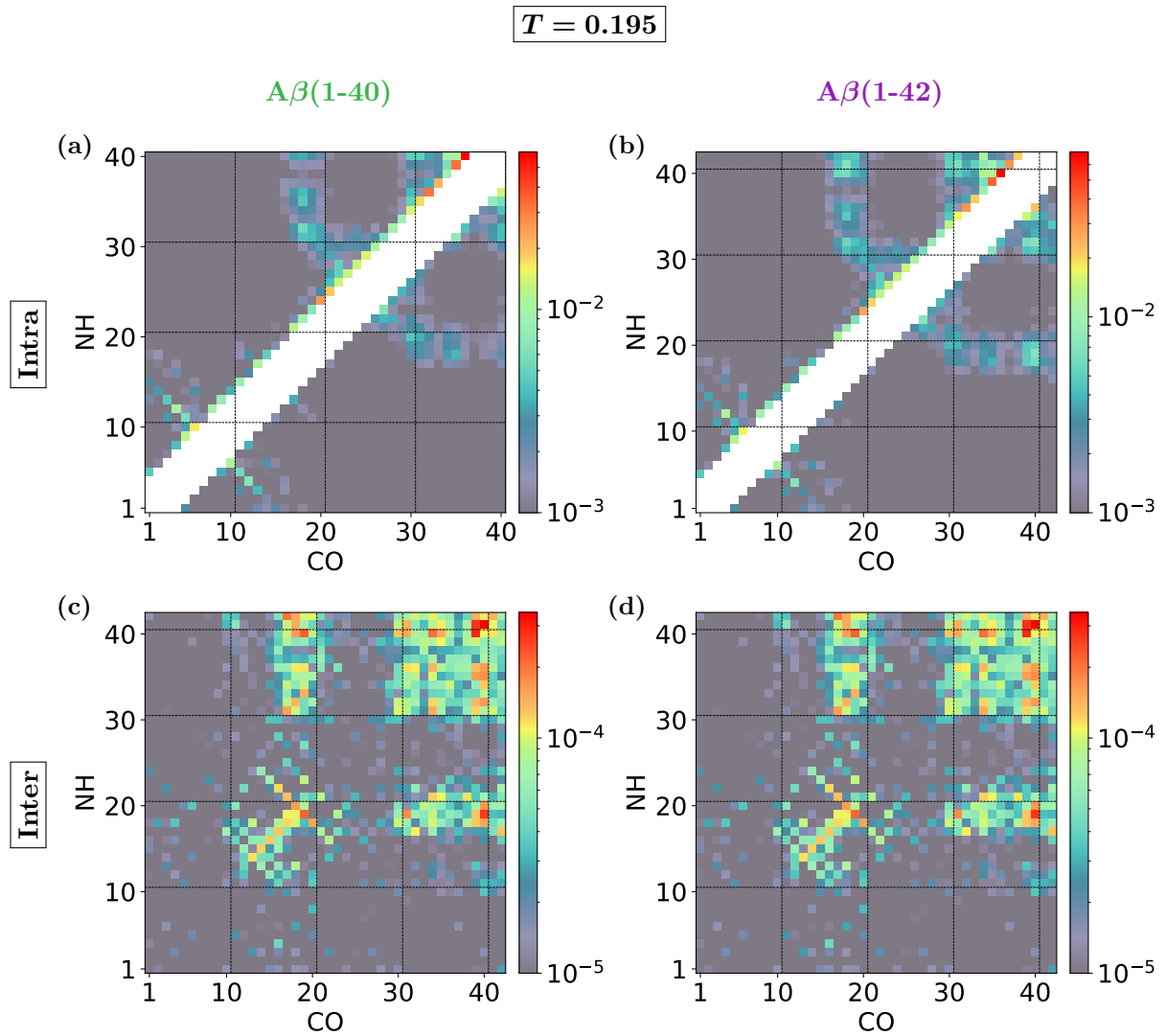
**Figure C.1:** Temperature and inverse temperature vs configurational energy for A $\beta$ (1-40) and A $\beta$ (1-42).



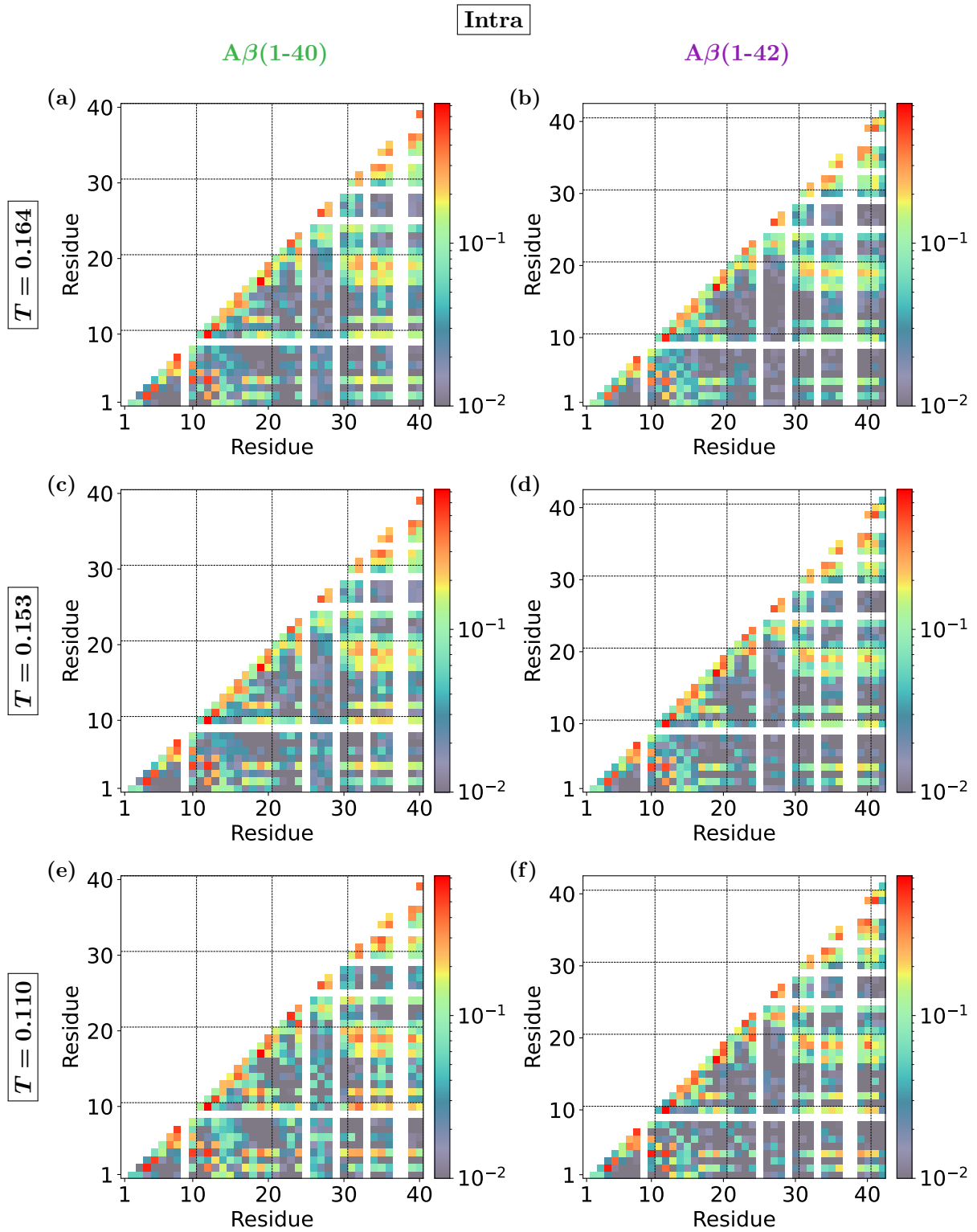
**Figure C.2:** Amyloid beta hydrogen bond contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$  with a non-logarithmic probability scale.



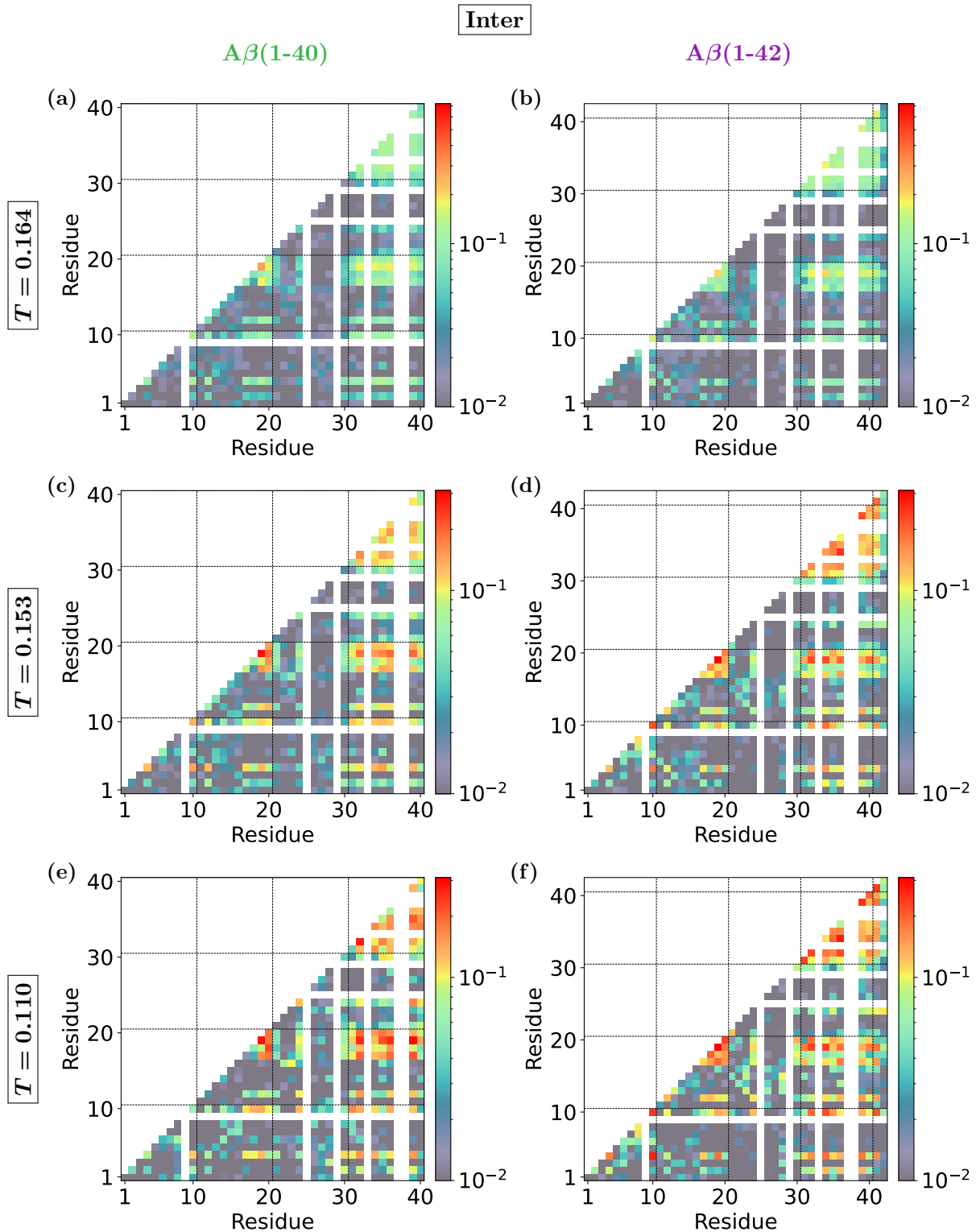
**Figure C.3:** Amyloid beta hydrogen bond contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$  with a logarithmic probability scale.



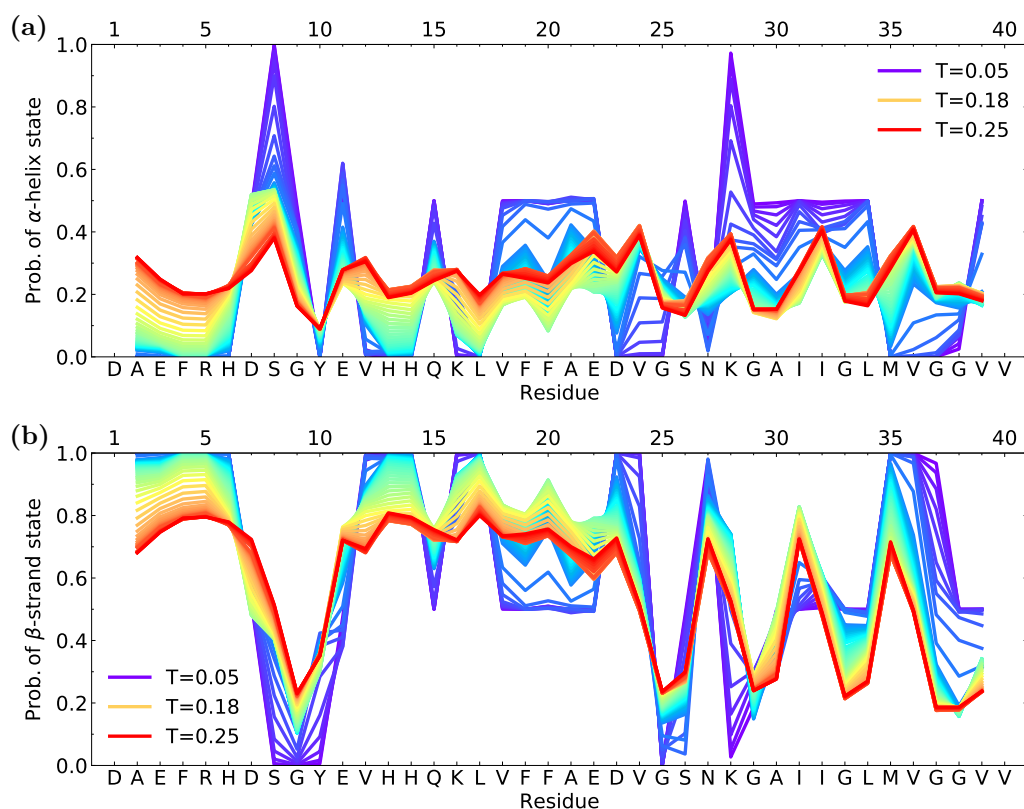
**Figure C.4:** Amyloid beta intra-molecular hydrogen bond contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$  in the unfolded state at  $T = 0.195$ .



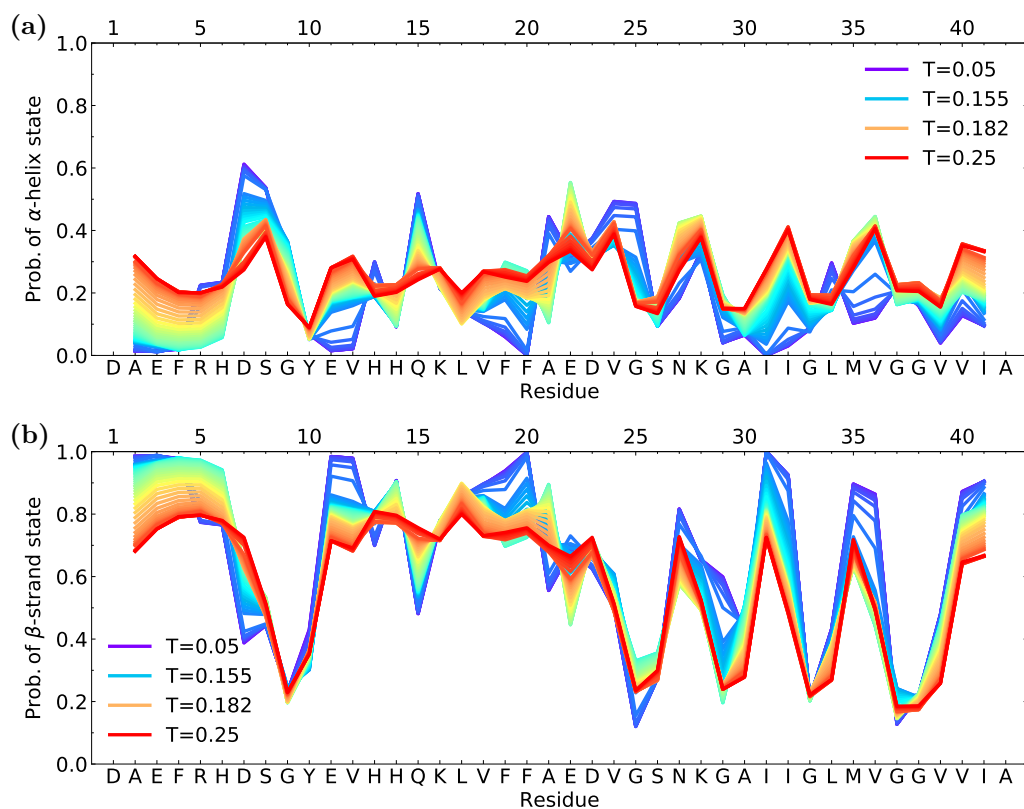
**Figure C.5:** Amyloid beta **intra**-molecular side chain contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$ .



**Figure C.6:** Amyloid beta **inter**-molecular side chain contact probability maps for  $A\beta(1-40)$  and  $A\beta(1-42)$ .



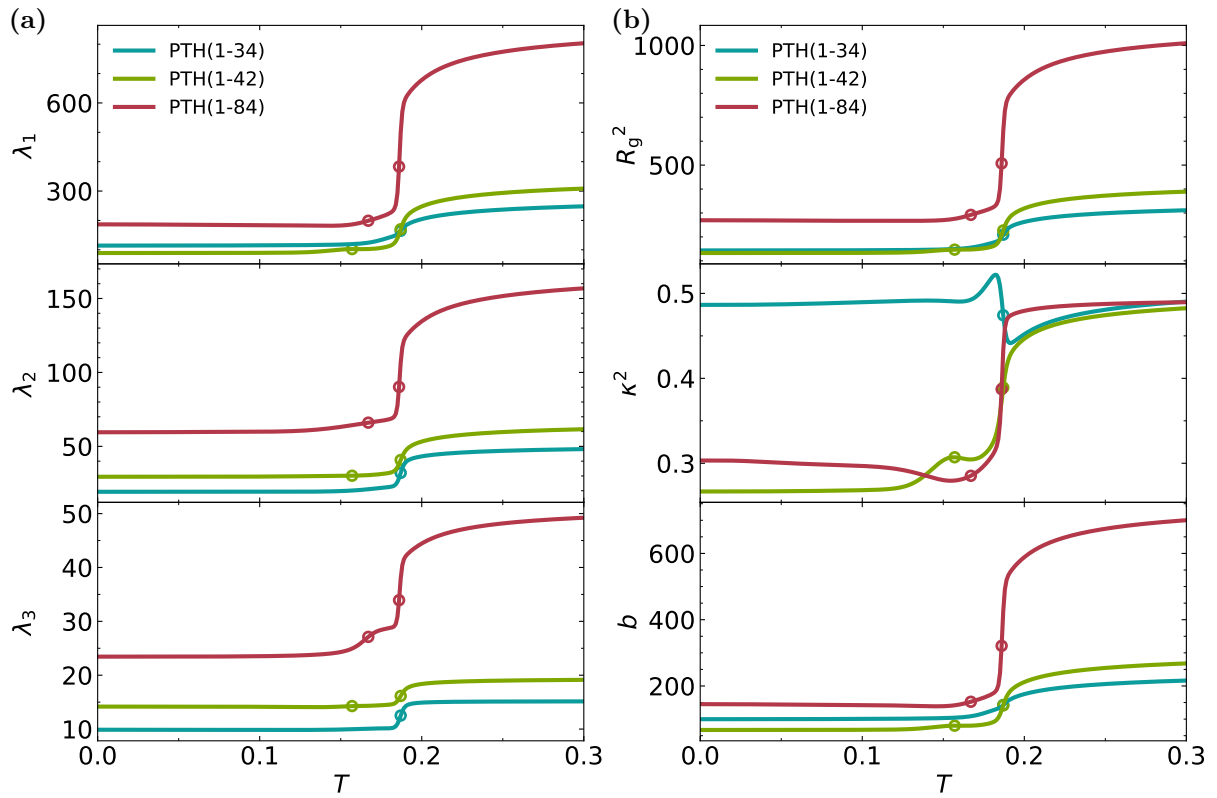
**Figure C.7:** Amyloid beta Ramachandran analysis for A $\beta$ (1-40).



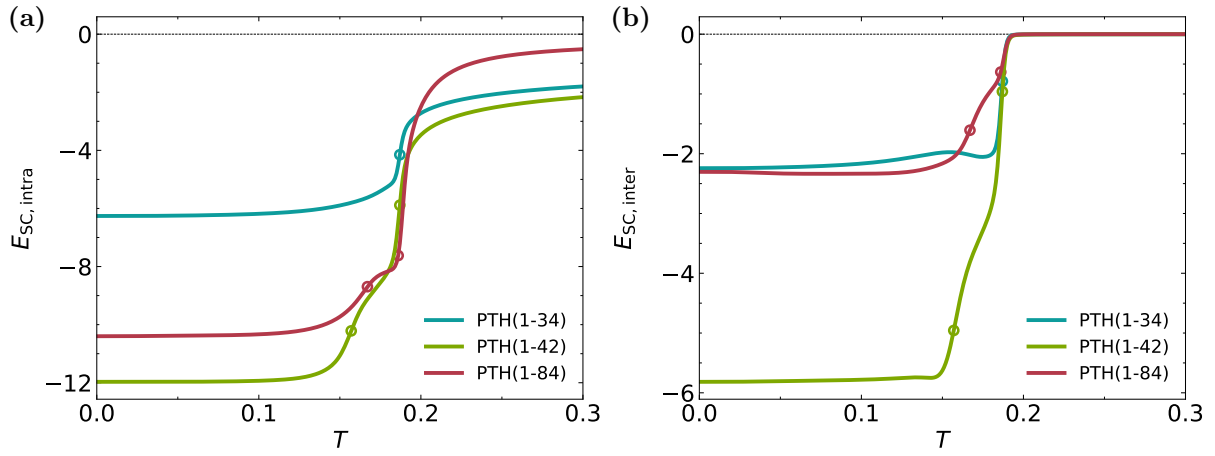
**Figure C.8:** Amyloid beta Ramachandran analysis for A $\beta$ (1-42).

## Appendix D

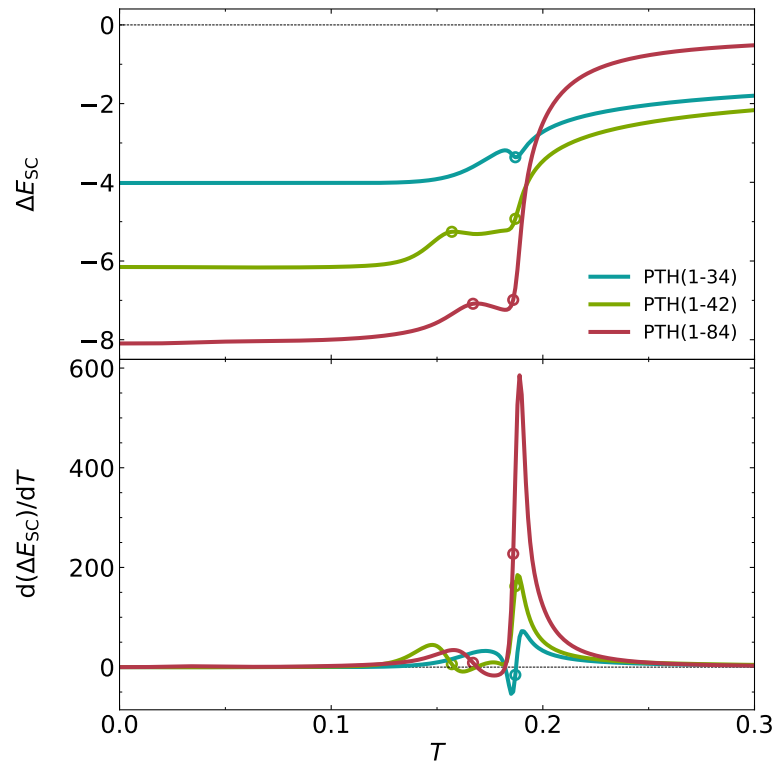
# Additional Figures for the Parathyroid Hormone Systems



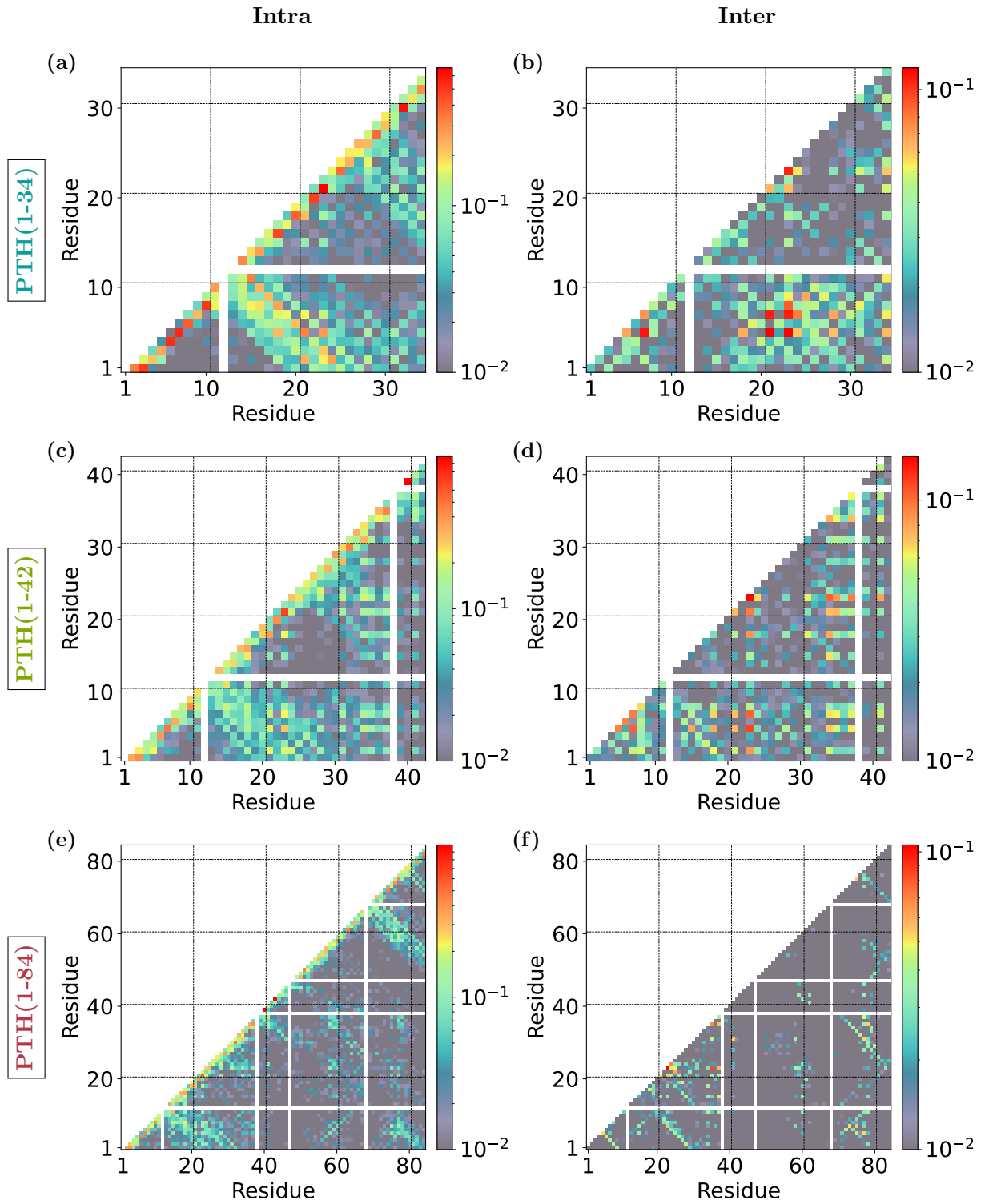
**Figure D.1:** PTH tensor of gyration. Eigenvalues and other derived properties.



**Figure D.2:** PTH intra- and inter-molecular side chain energies.



**Figure D.3:** Difference between intra and inter-molecular side chain energies.



**Figure D.4:** PTH intra-molecular and inter-molecular side chain contact probability maps for  $T = 0.178$ .

# Acknowledgement

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Wolfgang Paul, my supervisor, for his invaluable guidance and mentorship throughout my studies. His support and encouragement introduced me to this fascinating field of research and ensured that my work stayed focused and meaningful. I am especially grateful for his openness and availability, as I could always approach him to discuss scientific problems. Working under his supervision has been an enriching experience.

I would also like to thank Prof. Dr. Daniel Sebastiani, who served as my mentor. Although our meetings were less frequent, his insightful comments on my work always proved invaluable, offering new perspectives on various topics. Additionally, I would be remiss not to mention that, in my opinion, he serves the best coffee in the institute!

My heartfelt thanks go to Dr. Timur Shakirov, who has been a constant source of knowledge and inspiration during this journey. As my "guru of SAMC", his unmatched expertise and experience have been instrumental in my progress. I particularly appreciated those daily routine moments when he would walk by my desk, glance at my work, and provide a single comment that would often spark new ideas or solutions.

I am also deeply grateful to Dr. Arne Böker, my predecessor on this project, who helped me get my bearings in what was initially a new and challenging topic. His foundational work on applying PRIME20 models to SAMC paved the way for my research. Beyond his academic contributions, I will always cherish his humor and patience, which made collaboration with him a joy.

Special thanks are due to Dr. Viktor Ivanov, who has been like a second mentor to me. His expertise in computational polymer physics and his constructive comments on my thesis greatly enriched my work. I am also thankful for the many fruitful discussions we had over the years.

To Dr. Lama Tannoury, my desk neighbor, I am grateful for her support and camaraderie. Whether it was technical issues or personal grievances, she was always there to lend a helping hand. Her presence made coming to the office a genuinely enjoyable experience.

I would also like to extend my heartfelt gratitude to Dr. Oliver Busch, who has been a constant source of humor, support and advice. I fondly remember our shared coffee breaks, outside of the institute building. Moreover, it is always important to have someone nearby, who can effortlessly quote from the "Känguru Manifest".

I would also like to acknowledge all the current and past members of our research group, including Prof. Steffen Trimper, Dr. Eunsang Lee, Dr. Benno Werlich, Dr. Michael Beyer, Kai-Hendrik Henk, and Philipp Tesch, for creating a collaborative and stimulating research environment.

A special thank you goes to Dr. Ann-Kristin Flieger, coordinator of the graduate school, for helping me settle into PhD life and for organizing a wide range of courses on scientific and science-adjacent topics. Her efforts were particularly appreciated during the challenging times of COVID-19 lockdowns. As a PhD student representative in the SFB, I had the privilege of working closely with her, and her dedication was inspiring.

I am also grateful to all the PhD students from the third funding period of the SFB TRR-102, particularly those who were part of the organizing committee of the "best conference that never happened": Stephan Sydow, Anna-Maria Tsirigoni, Wing Kit Or, André Paschold, Yu

Qiang, Anika Wurl, Yue Cai, Twinkle Bhatia, and Thomas Kunze.

I am sincerely thankful to the DFG for funding me through the SFB TRR-102 project, which provided me with the resources and financial security to focus on my research.

On a more personal note, I wish to express my heartfelt gratitude to my parents for their unwavering support. Their financial help and encouragement have been vital to my success, and their upbringing has provided me with the tools and resilience to reach this point.

To my wife, Dana Sommerfeld, I am endlessly thankful for her love, patience, and support. She has been my rock, listening to my countless rants, rambles, and tears about computational polymer physics, and I could not have done this without her.

I am also grateful to my brother, Matthias, and to my many friends whose love and support have given me strength throughout this journey. In no particular order, I want to acknowledge Vincent Streichhahn, Anika Zorn, Tom Göhring, Lena Bonkat, Felix Böhmer, Ezra Church, Johannes Banholzer, Julian Bauch, Nicole Tröger, Sanja Ude, Marc Münster, John Frenzel, Eleonora Heiden, and all the many others who make my life brighter and more fulfilling.

Lastly, to everyone who has been a part of my journey – whether named or unnamed – thank you for making this time in my life so meaningful.

# Angaben zur Person

Vorname	Christian
Nachname	Sommerfeld
Geburtsname	Lauer
Geburtsdatum	
Geburtsort	
Staatsangehörigkeit	
Geschlecht	
Anschrift	
vorhandener akademischer Grad	M. Sc.
Fachgebiet der Promotion	Theoretische Polymerphysik

## Bildungsgang

10/2012 – 03/2017	<b>Bachelor of Science in Physik</b> Martin-Luther-Universität Halle-Wittenberg Titel der Bachelorarbeit: <i>"Lennard-Jones particles in the Vicsek model"</i>
04/2017 – 06/2020	<b>Master of Science in Physik</b> Martin-Luther-Universität Halle-Wittenberg Titel der Masterarbeit: <i>"Thermodynamics of the Dimerization of Polyglutamine and Polyalanine chains: A Monte Carlo Simulation in the Prime20 model"</i>
seit 08/2020	<b>Promotionsstudium</b> Martin-Luther-Universität Halle-Wittenberg Fachgebiet: Theoretische Polymerphysik

---

Ort, Datum

---

Unterschrift



# Publikationsliste

- ▷ **C. Lauer**, W. Paul "Dimerization of Polyglutamine within the PRIME20 Model using Stochastic Approximation Monte Carlo"
- ▷ T. Kunze, C. Dreßler, **C. Lauer**, W. Paul, D. Sebastiani "Reverse Mapping of Coarse Grained Polyglutamine Conformations from PRIME20 Sampling". In: ChemPhysChem 25.9 (2024)
- ▷ T. Kunze, C. Dreßler, **C. Lauer**, D. Sebastiani "Assignment of a Physical Energy Scale for the Dimensionless Interaction Energies within the PRIME20 Peptide Model". In: ChemPhysChem 202400592 (2024), pp. 1–6



# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit mit dem Titel *Thermodynamics of the dimerization of amyloidogenic proteins: A computer simulation study using Stochastic Approximation Monte Carlo* selbständig und ohne fremde Hilfe angefertigt habe. Die verwendeten Quellen und Hilfsmittel sind vollständig angegeben. Die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen sind als solche kenntlich gemacht.

Ich erkläre, dass ich alle Angaben wahrheitsgemäß gemacht habe und bisher keine vergeblichen Promotionsversuche unternommen, und dass ich die Dissertation weder in der gegenwärtigen noch in einer anderen Fassung bereits einer anderen Prüfungsbehörde vorgelegt habe. Außerdem erkläre ich, dass ich mich noch bei keiner anderen Hochschule einem Promotionsverfahren unterzogen oder diese begonnen habe.

Ich bin weder vorbestraft, noch sind Ermittlungsverfahren gegen mich angehängt.

---

Ort, Datum

---

Unterschrift