June, 2015

Specificity and characteristics of fuzzy complexes

Modelling of intrinsically disordered proteins

Karolina Mothander

Department of Astronomy and Theoretical Physics, Lund University

Bachelor thesis supervised by Stefan Wallin

LUND UNIVERSITY

Abstract

Intrinsically disordered proteins are fully functional proteins without a three-dimensional structure. Intrinsically disordered proteins can, in the same way as ordered proteins, interact with other proteins to fulful their function. When intrinsically disordered proteins interact with other proteins they can form complexes, so-called fuzzy complexes, with different degrees of disorder still present. Fuzzy complexes are found in a number of different places, e.g. in transcription factors.

This thesis will try to answer the question, if and how specificity is affected by fuzziness, and how different characteristics affect fuzziness. This is done by a coarse-grained model with two protein chains, the target and the probe. The first step is a design process where the probe learns the target. The second step is a recognition step, where the probe is exposed to a number of different rival targets and will have to recognise the original target. The last step investigates the characteristics of the target. The design process is repeated for targets with different characteristics. The model uses ordered and fuzzy complexes that are modeled in two different cases, with only hydrophobic - polar (HP) amino acids and all with twenty amino acids (20x20), respectively.

It was found that fuzzy complexes can interact in a specific way, but they are less specific than ordered complexes. This means that fuzzy complexes can interact specifically with a number of different proteins. No clear trend was found between the characteristics of the target and the fuzziness of the probe.

Contents

List of Figures

1 Introduction

Proteins are essential to the function of a cell. They take part in translation of DNA and messaging in and between cells. Proteins consist of long chains with different amino acids. There are twenty amino acids with different characteristics; they can be hydrophobic or polar, and some are positively or negatively charged. The order of the amino acid chain is determined by the genetic code in the DNA, which also determines the structure of the protein. Locally, proteins can fold into α-helices and β-sheets, and globally into three-dimensional structures. Proteins also have the ability to interact with other proteins and form complexes. It is often through interaction with other proteins that proteins become functional. Through these complexes the proteins can carry out their function (Johnson et al., 2015).

For a long time it was believed that the global structure of a protein determined its function and that if a protein did not fold correctly that function was lost. With more proteins getting sequenced and their three-dimensional structure solved, more functional unfolded proteins are found. These proteins often have a low complexity in their sequence. Proteins with this unfolded structure are called intrinsically disordered proteins due to the nature of the protein structure (Wright and Dyson, 1999). A study preformed by Romero et al. (1998) suggests that as many as 15,000 proteins in the Swiss Protein Database have disordered regions and many of those are least 40 amino acids long.

Intrinsically disordered proteins can interact with other proteins, in a similar way as ordered proteins, to fulfil their function. The difference being that intrinsically disordered proteins do not have a specific three-dimensional structure, but can fold into a more or less structured structure when interacting with another protein. Ordered proteins that already have a three-dimensional structure will keep that structure when binding and will be less flexible than disordered proteins (Wright and Dyson, 2009). The different complexes formed when an intrinsically disordered protein interacts with other proteins can be classified according to the magnitude of disorder still present in the formed complex. Tompa and Fuxreiter (2008) suggest four categories from static to dynamic disorder, with the collective term fuzziness, for different disordered proteins. A question that arises from these observations is if, and how, protein interaction can remain specific when the complex is dynamic.

There is a number of different examples of fuzzy complexes and intrinsically disordered proteins. One is the Arf-Hdm3-p53 pathway, which regulates tumour suppres-

sion. Very little was previously known about this pathway due to the intrinsically disordered regions and large size of Arf and Hdm2. With new technique it was found that the disordered proteins Arf and Hdm2 fold when they interact (Sivakolundu et al., 2008).

Intrinsically disordered proteins are also common in transcription factors. For example, the transcription factor CREB has an intrinsically disordered region, pKID. pKID binds to the KIX region of the transcriptional co-activator CPB. When pKID binds to KIX it will undergo a transformation from disordered to an intermediate form, and pKID starts to attain its structured form. pKID and KIX will then form a complex where pKID is ordered into two α-helices. This process is an example of coupled folding and binding (Sugase et al., 2007). The pKID - KIX complex have been theoretically modeled by Turjanski et al. (2008). They used a so-called G$\bar{o}$-type model with each residue represented by a single bead. This model is used to look at folding and interaction between proteins, and is just one of the many models used for this purpose.

Another type of model is the so-called coarse-grained model, which also uses a single bead to represent the amino acid. The coarse-grained lattice model presented by Behringer et al. (2006) is the basis for the model developed in this thesis. The aim is to produce a general model for fuzzy complexes, i.e protein interaction between an intrinsically disordered protein and its target. The disorder is introduced through loops in the protein chain and entropy is used as a measure for fuzziness, higher entropy represents larger fuzziness. The model is first used in a simple hydrophobic - polar case, where the hydrophobic and polar residues are represented by 1 and -1. Energies for all of the twenty amino acids are then introduced. The model is aimed to show the nature of fuzzy complexes and how specificity is affected by fuzziness.

2 Models and methods

2.1 HP model for protein-protein interaction

The model developed in this thesis is based on a coarse-grained model (Behringer et al., 2006) and uses two protein sequences, the target, $\sigma = (\sigma_1, \ldots, \sigma_N)$, and the probe, $\theta = (\theta_1, \ldots, \theta_N)$. These sequences are thought of as the surface exposed parts of two proteins, whose interaction is to be studied. The probe and the target are kept short in order to simplify the simulations. In the hydrophobic-polar (HP) model, σ_i and θ_i can be either $+1$ or -1 corresponding to hydrophobic and polar residues,

respectively. The calculations consist of two steps. The first is a stochastic design step by which the probe learns to recognise a given target, σ. Each probe sequence, θ, is assigned a statistical weight, $P(\theta) \propto \exp(-\beta_{\mathrm{D}}\mathcal{H})$, where β_{D} is a inverse design temperature and $\mathcal{H}$ is given by

$$\mathcal{H} = -\epsilon \sum_{i=1}^{\mathrm{N}} \theta_i \sigma_i, \tag{2.1}$$

and thought of as an effective interaction energy. In the second step, the ability of the θ ensemble to distinguish between the given target and other targets is tested.

The θ ensemble is generated using a Monte Carlo simulation. The simulation starts by choosing a probe sequence at random and a starting energy is calculated for that state according to the Hamiltonian in eq. (2.1). In the next step θ is updated, to get a new state, by choosing a random position on θ and giving it a new value, in this case -1 will become 1 and 1 will become -1. A energy for the new state is then calculated and the state will be accepted with a certain probability which is dependent on the inverse design temperature, β_{D}, and the difference in energy between the new and the old state, $\mathrm{P}(\theta_{\mathrm{old}} \rightarrow \theta_{\mathrm{new}}) = \min\big(1, \exp\big[-\beta_{\mathrm{D}}\,\Delta\mathcal{H}(\theta, \sigma)\big]\big)$, where $\beta_{\mathrm{D}} = 1/k_B\mathrm{T}$. If the state is accepted, the process will start again with an update of θ and a new energy is calculated for the new state. If the state is rejected, θ is rejected and the step is redone. The algorithm is summarised in algorithm 1 and is implemented in C.

At the end of the design process an ensemble of probe molecules with different energies will have been created. The average energy, $\langle E \rangle$, for the ensemble of probe molecules is plotted against β_{D}. The complementarity between the probe and a target,

$$K = \sum_{i=1}^{\mathrm{N}} \sigma_i\, \theta_i, \tag{2.2}$$

is also calculated for the different β_{D}'s and the average complementarity, $\langle K \rangle$, for the θ ensemble is plotted against β_{D}.

The next step is to test the recognisability of the probe against other target molecules, so called rival target molecules, $\sigma^{(1)}$. The probe molecules are tested against the rival targets and the average energy $\big(\langle E^{(1)} \rangle\big)$ for the each of the rival targets are

Algorithm 1: Example of the Monte Carlo simulation used in the design process.

Start by randomly generating σ and θ

1. Choose random position, j, in θ
2. Update θ_j, $\theta_j \to -\theta_j$, to a new θ state
3. Calculate energy according to eq. (2.1)
4. Accept the new state with probability,
$\mathrm{P}(\theta_{\mathrm{old}} \to \theta_{\mathrm{new}}) = \min\big(1,\, \exp\big[-\beta_{\mathrm{D}}\, \Delta\mathcal{H}(\theta,\, \sigma)\big]\big)$
if *accepted* **then**
save the new θ state to the ensemble
else
add another copy of the old state to the ensemble
5. Go to step 1, using the last added state as the old state. (ten million iterations)

Returns θ ensemble

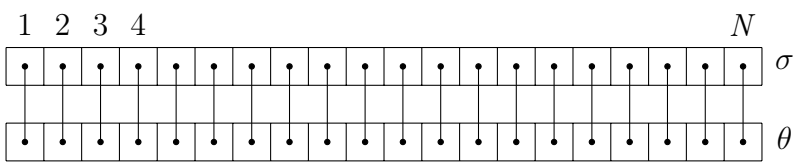

Figure 2.1: An example of ordered interaction. The probe and the target interact directly.

calculated. The difference (ΔE) in average energy between the target, $\sigma^{(0)}$ $\big(\langle E^{(0)}\rangle\big)$, and the rival target, $\sigma^{(1)}$,

$$\Delta E = \langle E^{(0)}\rangle - \langle E^{(1)}\rangle, \tag{2.3}$$

is used as a measure of recognisability and is calculated for the different rival targets. ΔE is plotted against the similarity, Q, between the target and rival target, defined as

$$Q = \sum_{i=1}^{\mathrm{N}} \sigma_i^{(0)} \sigma_i^{(1)}. \tag{2.4}$$

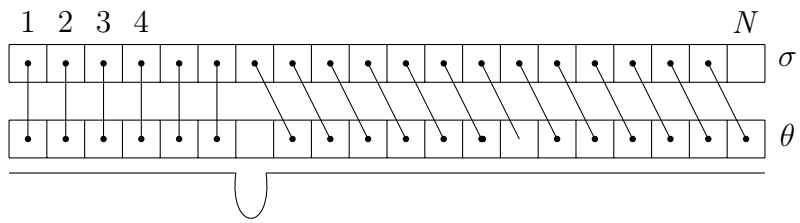

Figure 2.2: An example of a fuzzy complex, with a loop in the probe, θ, at amino acid seven. Each probe sequence is one unit longer then the target sequence. The residue in the loop does not interact with any target residue.

2.2 Extended HP model

So far the interaction between the probe and the target has looked like fig. 2.1, i.e. both the target and the probe have been ordered, and σ_1 has interacted with θ_1, σ_2 with θ_2, and so on. In order to introduce fuzzy interaction, N different energies are calculated in each step in the design process. Every energy corresponds to a loop at a probe amino acid, meaning that the probe is the intrinsically disordered protein chain. The first energy corresponds to a loop at the first θ position, the second energy to a loop at the second position, and so on, with the last energy corresponding to a loop at the last θ position. Figure 2.2 shows how the loops are introduced. In order to make the same types of calculations on the θ ensemble as in section 2.1 the free energy, F, is used, meaning that $\Delta\mathcal{H}$ in algorithm 1 will change to ΔF. The free energy is related to the N different energies in the following way,

$$F(\theta) = -\frac{1}{\beta_D} \ln \left[\mathrm{e}^{-E_1\beta_D} + \mathrm{e}^{-E_2\beta_D} + \cdots + \mathrm{e}^{-E_N\beta_D} \right]. \tag{2.5}$$

Free proteins have a high entropy and when they bind to other proteins the entropy will decrease. If instead a disordered protein binds to form a fuzzy complex the entropy will still be high. Entropy, defined as,

$$S(\theta) = -k_B \sum_k^{\mathrm{N}} P_k \ \ln \ P_k, \tag{2.6}$$

is therefore introduced as a measure of fuzziness. P_k is the probability of populating state k, $P_k(\theta) = Z^{-1}e^{-\beta E_k}$ where Z is the partition function, $Z(\theta) = \sum_k e^{-\beta E_k}$. Both the free energy and the entropy are calculated for different β_D's.

The ensemble averages of the free energy, ($\langle F \rangle$), and of the entropy ($\langle S \rangle$) are plotted against β_D. The recognisability of the fuzzy protein complexes is also tested, in the same way as for the non fuzzy complexes.

The last part will examine how the characteristics of the target, σ, will affect the fuzziness of the probe, θ. The design process is repeated with different σ, i.e. σ with high or low hydrophobicity. For each new σ an ensemble of θ are produced. The average free energy and average entropy for the ensemble are calculated and the entropy is plotted against the number of hydrophobic amino acids in σ.

2.3 20x20 model

In order to make the model more realistic, energies from all the twenty different amino acids are introduced. The energies used are from Lu et al. (2003). Similar to the HP model, a design process is performed where the probe learns the target (see section 2.1). Energies are calculated at the different inverse design temperatures, β_{D}. Fuzzy interaction is introduced as in section 2.2, and entropies and energies are calculated for different inverse design temperatures. The difference is that the Hamiltonian is represented by the energy matrix (M, Lu et al. 2003) for the twenty amino acids,

$$\mathcal{H} = \epsilon \sum_{i=1}^{\mathrm{N}} \mathrm{M}(\sigma_i,\, \theta_i). \tag{2.7}$$

The next step, which is the recognisability of the θ ensemble, is done in the same manner as in the HP model, in section 2.1. Here, the similarity parameter, Q, will take into account which amino acids have a polar or hydrophobic residue.

$$Q = \sum_{i=1}^{\mathrm{N}} m\big(\sigma_i^{(0)},\, \sigma_i^{(1)}\big), \tag{2.8}$$

where,

$$m\big(\sigma_i^{(0)},\, \sigma_i^{(1)}\big) = \begin{cases} 1, & \text{if both hydrophobic or both polar,} \\ 0, & \text{otherwise.} \end{cases}$$

The recognisability of θ, both for the fuzzy and non fuzzy complexes, was plotted against the similarity, Q, between the target, $\sigma^{(0)}$, and the rival target, $\sigma^{(1)}$. The characteristics of σ were also examined in the 20x20 model. In addition to the hydrophobicity, the total and net charge of the target, σ, was included.

3 Result

3.1 HP model for protein-protein interaction

Initially, the properties of the original HP model of Behringer et al. (2006) were examined. The model concerns protein-protein interaction in an ordered protein complex. The results attained here are very similar to those attained by Behringer et al. (2006). Figure 3.1 shows the average energy in an ordered protein complex plotted against the different inverse design temperatures, β_{D}. The minimum value reached at high β_{D} is due to the Hamiltonian in eq. (2.1) and that only polar ($= 1$) and hydrophobic ($= -1$) amino acids are used. Behringer et al. (2006) also tested the complementarity between the probe, θ, and the target, σ. Figure 3.2 shows the complementarity plotted against the β_{D}. The complementarity is calculated using eq. (2.2) and is normalised to get the complementarity per site. Figure 3.2 shows that the complementarity increases with increasing β_{D}, and at high β_{D} the probe and the target are identical.

3.2 Extended HP model

The next step was to introduce fuzziness to the model, in order to investigate the properties of fuzzy complexes. Fuzziness was introduced through loops in the probe, θ (see fig. 2.2). A first step was to see how the free energy of the fuzzy complex varied with β_{D}. Figure 3.3 shows the average free energy against β_{D}, and the free energy is given by eq. (2.5), which is the free energy only for the bound states. At low β_{D} the temperature goes towards infinity ($\beta_{\mathrm{D}} \propto 1/T$) and a bound state at infinite temperature is unrealistic, since the system could not be bound at those temperatures. Furthermore, for high temperatures, the entropy is underestimated since the number of bound states is limited to only one loop per amino acid, in reality there would be more bound states. The next property to be examined is how the fuzziness of the probe is affected by the inverse design temperature. Entropy is used as a measure of fuzziness and is given by eq. (2.6). Figure 3.4 shows the average entropy at the different β_{D}. The figure shows that the entropy goes towards three at low β_{D}, in

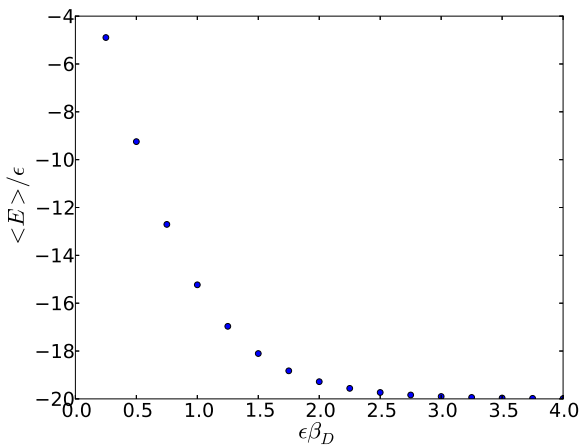

Figure 3.1: HP model. Average energy for a non fuzzy complex plotted against β_{D}. The energy is given by eq. (2.1).

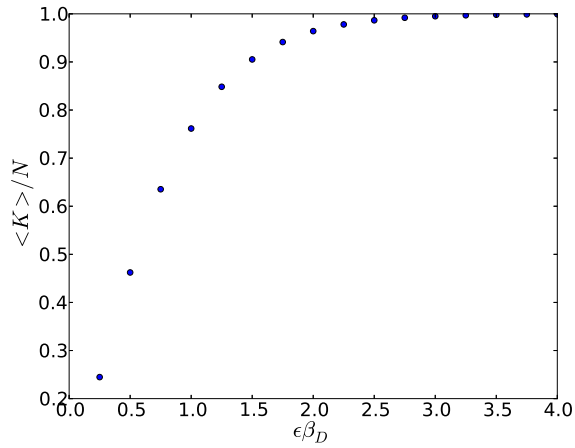

Figure 3.2: HP model. Average complementarity per site against β_{D}. The complementarity is given by eq. (2.2). Large values of $< K > /N$ means that the probe and the target are the same.

agreement with $\ln N = \ln 20 \approx 3$.

At low β_{D} the probe will be chosen almost at random due to the nature of the design process. This means that there is an equal probability for each of the states k in eq. (2.6). This means that $P_k = 1/N$, and eq. (2.6) gives,

$$S(\theta) \approx -N \cdot \frac{1}{N} \ln \frac{1}{N} \ln N. \tag{3.1}$$

This means that the entropy scales as $\ln N$, which is due to that only a limited number of bound states is used in the simulations. If loops at more than one amino acid were allowed that might push the entropy to scale with N instead.

One of the main questions in this thesis is how the specificity of protein interaction is affected by protein disorder. The specificity is tested by taking the probe molecules from the design processes, test them against rival targets, and see if the probe molecules recognise the target instead of the rival target. The specificity or recognisability was tested for both ordered and disordered proteins, in order to compare the recognisability of ordered and fuzzy complexes. Figure 3.5 shows the recognisability of the fuzzy complex (circles) and the ordered protein complex (triangles) as a function of the similarity between the target, $\sigma^{(0)}$, and the rival target,

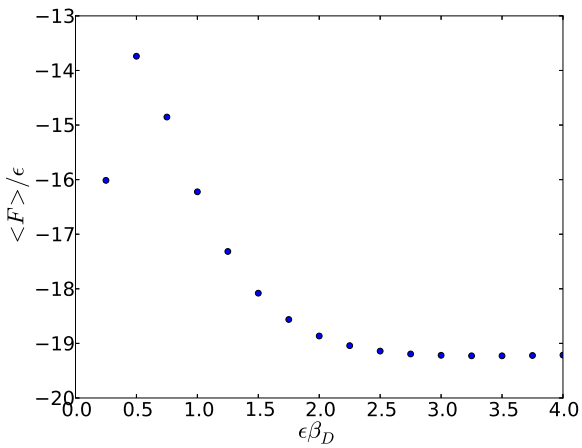

Figure 3.3: HP model. The average free energy of the fuzzy complex against the inverse design temperature, β_{D}. The free energy is given by eq. (2.5).

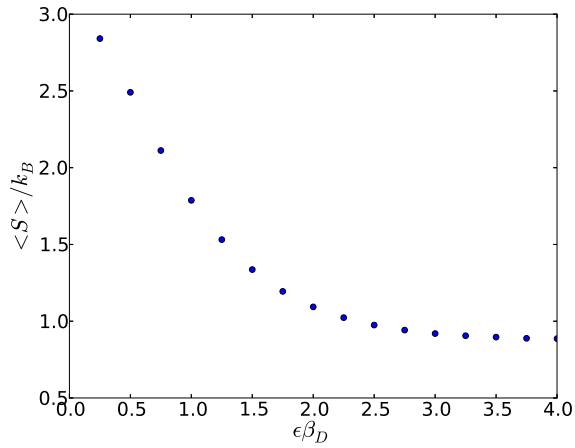

Figure 3.4: HP model. The average entropy of the fuzzy complex against the inverse design temperature, β_{D}. The entropy is given by eq. (2.6).

$\sigma^{(1)}$. The similarity, Q, is calculated using eq. (2.4), and is normalised to get the similarity per site. Zero means that half of the sites in the target and the rival target are the same. ΔF is calculated as the difference between the free energy for the target and the free energy for the rival target, $\Delta F = \left\langle F^{(0)} \right\rangle - \left\langle F^{(1)} \right\rangle$. The coefficient of determination, R^2, was calculated for the fitted line in this plot and for the fitted lines in subsequent plots. The R^2 is defined as

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}}, \tag{3.2}$$

where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. The R^2 is equal to 1.000 and 0.827, for the ordered complex and the fuzzy complex, respectively. A negative ΔF means that the target is prefered instead of the rival target.

The last thing that was investigated was how the characteristics of the target, σ, affects the fuzziness of the probe, θ. The design process was redone with a number of different targets with different amounts of hydrophobicity. The entropy (eq. (2.6)) was calculated for each target. Figure 3.6 shows the average entropy as a function of the number of hydrophobic amino acids, i.e. how many are 1 or -1, in the target. The dashed line is an average of the data points at a specific hydrophobicity, and the

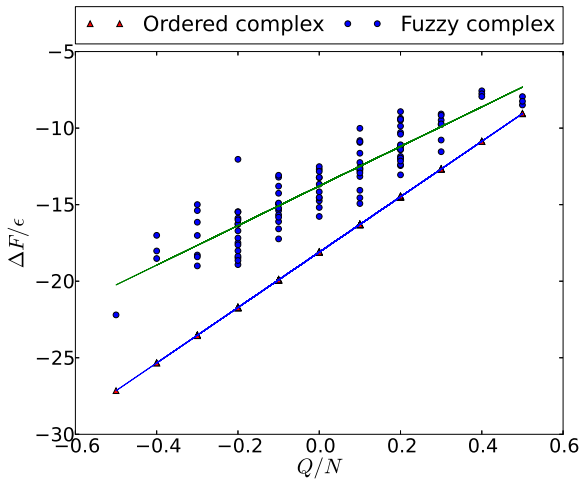

FIGURE 3.5: The difference between the energy of the target and the energy of the rival target. Circles are the fuzzy complex and triangles are the ordered complex. The R^2 for the non fuzzy complex is 1.000 and the R^2 for the fuzzy complex is 0.827.

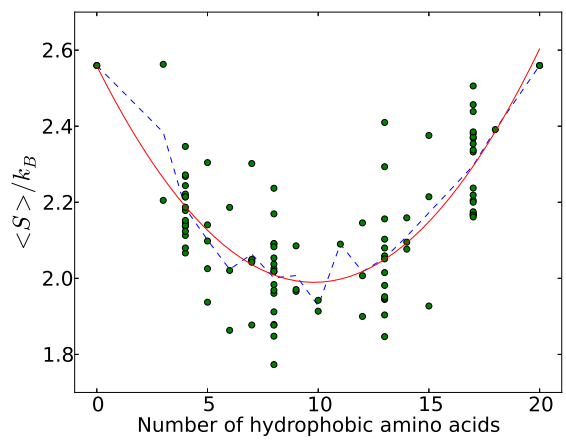

FIGURE 3.6: Characteristics of the target, σ. The number of hydrophobic amino acids in σ is on the x-axis and the entropy is on the y-axis. The dashed line is an average of the entropy at a specific hydrophobicity. The solid line is a second degree polynomial.

solid line is a fitted second degree polynomial. The target is fuzziest at low or high numbers of hydrophobic amino acids in the target. When all the amino acids in the target are the same, that would lead to the probe amino acids to be the same. This would lead to a high probability of the energies to populate any states k in eq. (2.6), which gives a high entropy.

3.3 20x20-model

This step represents the final extension of the model, where energies from all the twenty amino acids are used. This extension follows similar steps as the previous sections. First, the energy of the ordered protein complex and the free energy of the fuzzy complex are investigated. Figure 3.7 shows the energy for the ordered protein complex, with energy given by eq. (2.7). Figure 3.8 shows the free energy for the fuzzy complex, the free energy is given by eq. (2.5) with the eq. (2.7) as the Hamiltonian. Similar trends as in the HP model are observed, the energy for the ordered complex converges towards a minimum value at high inverse design temperatures, β_{D}, and the free energy follows the same trend as in fig. 3.3.

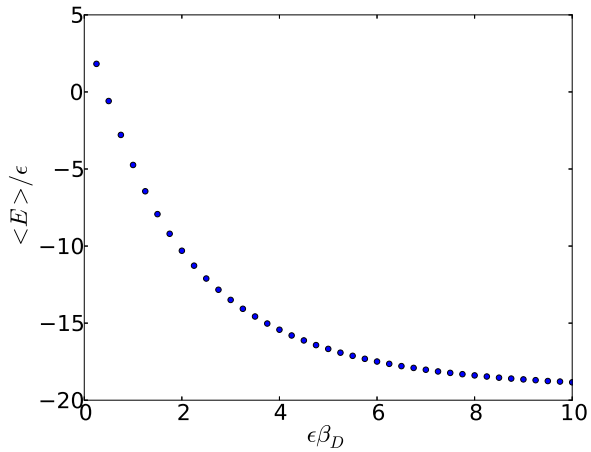

FIGURE 3.7: 20x20 model. Average energy for a non fuzzy complex against the inverse design temperature, β_{D}. The energy is given by eq. (2.7).

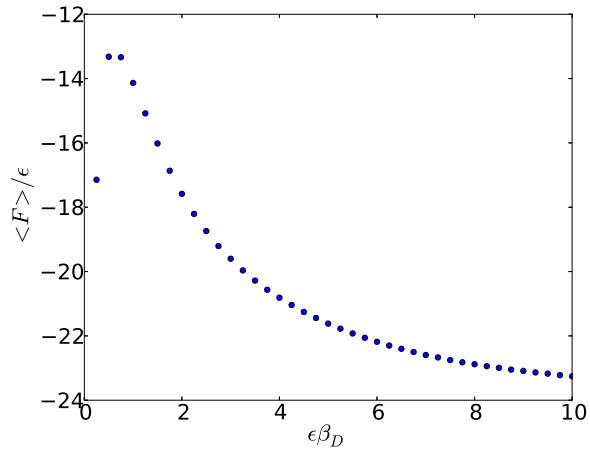

FIGURE 3.8: 20x20 model. Average free energy for a fuzzy complex against the inverse design temperature, β_{D}.

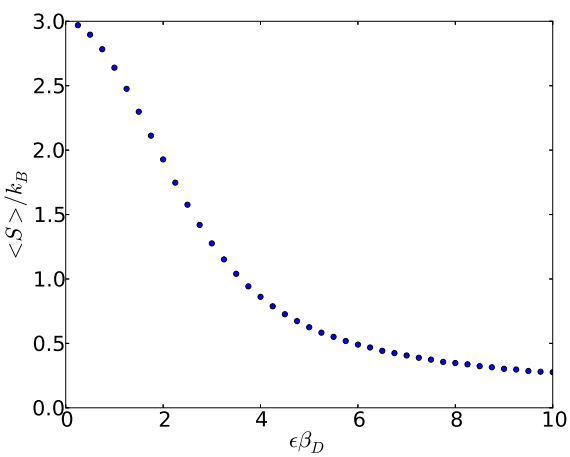

FIGURE 3.9: 20x20 model. Average entropy against the inverse design temperature, β_{D}. The entropy is given by eq. (2.6).

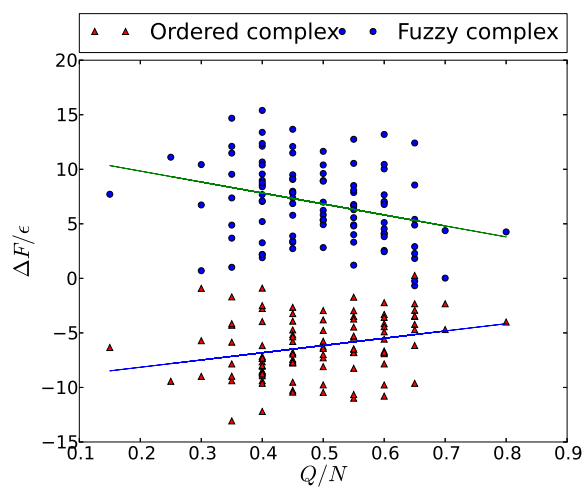

FIGURE 3.10: 20x20 model. Recognisability of the fuzzy (circles) and ordered complexes (triangles). R^2-values are 0.097 and 0.075 for the fuzzy and ordered complexes, respectively. Q/N is normalised to give similarity per site and Q is given by eq. (2.8), $Q/N = 1$ represents full similarity.

The amount of fuzziness as a function of inverse design temperature is also examined in the 20x20 model. Entropy (eq. (2.6)) is also used in this case and fig. 3.9 shows the entropy as a function of β_{D}. It is clear, even in this case, that the entropy reaches a maximum at three, which is due to the number of amino acids used in the model.

The specificity of the ordered and fuzzy complexes is also investigated in the 20x20 model. Figure 3.10 shows the recognisability for the fuzzy (circles) and ordered (triangles) complexes. Similar to the HP model, ΔF is the difference in energy between the target and rival target. In the 20x20 model the similarity, Q, is defined according to eq. (2.8) which takes into account if the amino acids in the target and the rival target are hydrophobic or polar. Q is then normalised to give the similarity per site, and Q/N ranges between zero and one, where $Q/N = 1$ represents complete similarity between the target and the rival target. R^2-values are 0.097 and 0.075 for the fuzzy and the ordered complexes, respectively. What is evident for the fuzzy complex is that $\Delta F > 0$, which means that the probe would prefer to bind to the rival targets instead of the target it was designed for.

The last property to be examined in the 20x20 model, is how the characteristics of the target, σ affects the fuzziness of the probe, θ. In the 20x20 model three characteristics are examined, the hydrophobicity, the net charge and the total charge of the target. Figure 3.11a shows the free energy between the probe and the target as a function of the hydrophobicity of the target. Figures 3.11b to 3.11d show the entropy as a function of hydrophobicity, net charge and totalt charge, respectively. The R^2-value is 0.772 for the free energy against the hydrophobicity (fig. 3.11a). The R^2-value is 0.522 for the entropy against the hydrophobicity (fig. 3.11b). The R^2-values are 0.059 and 0.302 for the net and total charge, respectively.

4 Discussion and conclusions

In this thesis we have extended a previous model (Behringer et al., 2006) for protein-protein interaction, in two ways. Fuzziness was added as a first extension, i.e interaction with intrinsically disordered proteins. Fuzziness was added through loops in the the probe amino acid chain and entropy was introduced as a measure of fuzziness. The second extension included the use of energies from all twenty amino acids, in order to make the model more realistic.

We have used the model to investigate the properties of fuzzy complexes and tried to

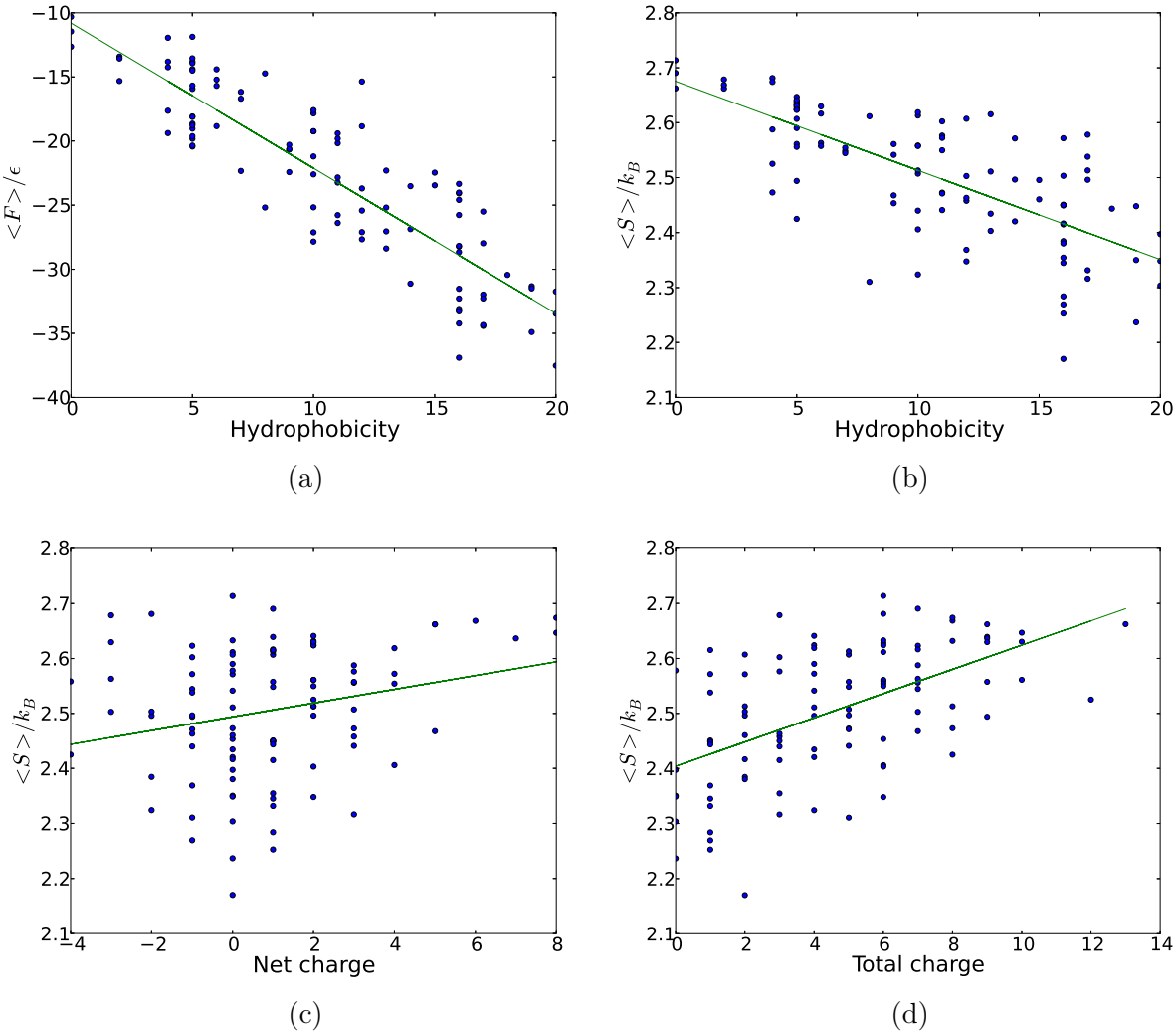

FIGURE 3.11: 20x20 model. The figures show how the characteristics of the target, σ, affect the probe, θ. (a) The average free energy as a function of the hydrophobicity, with $R^2 = 0.772$. The average entropy (b) as a function of the hydrophobicity, with $R^2 = 0.522$, (c) as a function of the net charge, with $R^2 = 0.059$, and (d)as a function of the total charge, with $R^2 = 0.302$.

answer the question of if and how disordered proteins can keep their specificity when they are dynamic. The model starts with a design process where a probe, θ, learns to recognise a target, σ. Due to the nature of the design process the energy have to decrease when β_{D} increses which is evident in fig. 3.1. The energy decreases when the similarity between the probe and the target increases, which is seen in fig. 3.2.

One of the main questions in this thesis is what happens to the specificity in a fuzzy complex. The specificity of the probe from the design process was tested by exposing the probe to a number of different rival targets, $\sigma^{(1)}$, to see if the probe could recognise the target, $\sigma^{(0)}$. The recognisability was tested for both the ordered and fuzzy complex in order to compare the two complexes. The target is recognised if the energy difference between the target and the rival target, $\Delta F = \left\langle F^{(0)} \right\rangle - \left\langle F^{(1)} \right\rangle$, is negative. If there is no difference in the energy then all the rival targets are recognised to the same extent as the target. The recognisability for the ordered and fuzzy complexes for the HP model are shown in fig. 3.5. The figure shows that the ordered complex shows a higher recognisability than the fuzzy complex. It also shows that both the fuzzy and ordered complexes have a certain specificity, the recognisability increases with an increase in similarity between the target and the rival target. The specificity is stronger for the ordered complex, which has a lower ΔF for all the data points than the ΔF for the fuzzy complex.

The recognisability was also tested with energies for the twenty amino acids, in the same way as in the HP model. The trend is more diffuse in this case, $R^2 = 0.075$ for the ordered complex in the 20x20 model, compared to $R^2 = 1.000$ for the ordered complex in the HP model, and $R^2 = 0.097$ for the fuzzy complex in the 20x20 model, compared to $R^2 = 0.827$ for the fuzzy complex in the HP model (see fig. 3.10). What is also evident in fig. 3.10 is that a majority of ΔF for the fuzzy complex is positive, indicating that the fuzzy complex prefers to bind to the rival target instead of the target it was designed for. One explanation can be that the hydrophobicity of the target affects the binding energy between the probe and the target, which can be seen in fig. 3.11a. The binding energy decreases with increasing hydrophobicity. The target used throughout the model had a hydrophobicity of six, which gives it a relatively high binding energy. The rival targets that the probe was tested against had a hydrophobicity between three and 15, giving some of the rival targets a lower binding energy which can explain why ΔF for the fuzzy complex is positive.

The last thing investigated in the model was if there is a connection between the characteristics of the target, σ, and the fuzziness of the probe, θ. The characteristics

of the target were examined by repeating the design process at a set inverse design temperature, β_{D}, but with different targets. A whole ensemble of probe molecules was produced for each new target and the entropy, fuzziness, were calculated for each target. The characteristics were tested in the HP model (see fig. 3.6) and in the 20x20 model (see figs. 3.11b to 3.11d). In the HP model it is possible to see a trend between the number of hydrophobic amino acids in the target and the fuzziness of the probe. High and low numbers of hydrophobic amino acids will give the largest fuzziness. In the 20x20 model there is no such clear trend, with an $R^2 = 0.522$ when the hydrophobicity is tested (fig. 3.11b), $R^2 = 0.059$ when the net charge is tested (fig. 3.11c), and $R^2 = 0.302$ when the total charge is tested (fig. 3.11d).

The reason no clear trend is seen in figs. 3.11b to 3.11d and only a weak trend is seen in fig. 3.6 can be due the simplicity of the model and that a chain of only 20 amino acids is used. The maximum entropy is affected by the number of amino acids used in the model. Trends can be made more visible with a longer amino acid chain. In the 20x20 model, where no trends can be seen, it can be a question of whether the right type of energy matrix was used for this kind of modeling. The energy matrix from Lu et al. (2003) is just one of many energy matrices for protein-protein interaction. Another possibility is that there is no connection between the fuzziness of the probe and the characteristics of the target.

To go back to the main question, is there still specificity in fuzzy complexes? In the HP model the answer is yes, there are clear trends that both the fuzzy and the ordered protein complexes have a specificity. The fuzzy complex is a bit less specific than the ordered complex, which is expected (Wright and Dyson, 2009). In the 20x20 model the case is not as simple. Figure 3.10 suggests that there is no or reversed specificity for the fuzzy complex. The trend seen can in part be explained with the decrease in binding energy with increasing hydrophobicity, as discussed above. However, there is also a weaker trend for the ordered complex, which suggests that there is something else affecting the recognisability. This can be due to the simplicity of the model. This is why it might be useful to extend the model further. First, to use a more hydrophobic target, with maybe around half or more hydrophobic amino acids. Second, to use a longer amino acid chain and allow the probe to be fuzzy in a different way, e.g. with loops at more than one amino acids. With a further extension of the model the trend might become clearer.

Acknowledgments

To Stefan Wallin for the help and support throughout the project. To Jacob Wikmark for advice with the code and proofreading of the finished report.

References

Behringer, H., Degenhard, A., and Schmid, F. Coarse-grained lattice model for molecular recognition. *Physical review letters*, 97(12):128101, 2006.

Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. *Molecular biology of the Cell*. New York : Garland Science, cop., 2015.

Lu, H., Lu, L., and Skolnick, J. Development of unified statistical potentials describing protein-protein interactions. *Biophysical journal*, 84(3):1895–1901, 2003.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. Thousands of proteins likely to have long disordered regions. In *Pacifc Symposium on Biocomputing (PSB)*, volume 3, pages 437–448, 1998.

Sivakolundu, S. G., Nourse, A., Moshiach, S., Bothner, B., Ashley, C., Satumba, J., Lahti, J., and Kriwacki, R. W. Intrinsically unstructured domains of arf and hdm2 form bimolecular oligomeric structures in vitro and in vivo. *Journal of molecular biology*, 384(1):240–254, 2008.

Sugase, K., Dyson, H. J., and Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–1025, 2007.

Tompa, P. and Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in biochemical sciences*, 33(1):2–8, 2008.

Turjanski, A. G., Gutkind, J. S., Best, R. B., and Hummer, G. Binding-induced folding of a natively unstructured transcription factor. *PLoS computational biology*, 4(4):e1000060, 2008.

Wright, P. E. and Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, 293(2):321–331, 1999.

Wright, P. E. and Dyson, H. J. Linking folding and binding. *Current opinion in structural biology*, 19(1):31–38, 2009.