



UNIVERSITY OF  
LIVERPOOL

**Identification and optimisation of ligands to  
target protein-protein interactions  
*EB1-SxIP proteins***

Thesis submitted in accordance with the requirements of the University  
of Liverpool for the Degree of Doctor of Philosophy

**Teresa Baptista de Almeida**

November 2016



## **ACKNOWLEDGEMENTS**

Firstly, I would like to express my sincere gratitude to my supervisors Dr Neil Berry and Dr Igor Barsukov for the continuous support during my PhD and related research for their patience (for my continuous doubts), their optimism (to keep my motivation), and their immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better supervisors. On the same note I would like to thank Dr Andrew Carnell for his support.

Besides my supervisors, I would like to thank Professor Lu-Yun Lian for her insightful comments and encouragement, but also for the hard questions and discussions that incentivised me to widen my research and see it from different perspectives.

My sincere thanks also go to Dr Thomas Zacharchenko and Dr Marie Phelan, who were almost secondary supervisors and taught me almost everything, since molecular biology, to protein expression and purification to protein NMR assignment and basic Linux. Without their precious support, it would not be possible to conduct this research.

I thank my fellow labmates for helping me with my day-to-day work, by answering questions and making a friendly working environment. I will not mention anyone in special as you all had an important part in this project.

Thanks to my friends, old and new, but especially the ones that made my life in Liverpool happier. Thanks for all the dinners, trips, drinks and most of all, your friendship.

Last but not the least, I would like to thank my family: especially my parents and my sister for supporting me throughout my life in general, listening to my complaints and always having something nice to say during difficult times. Without your support, I would not have made it.



## ABSTRACT

End binding protein 1 (EB1) is a key element in the complex network of protein-protein interactions at microtubule growing ends which has a fundamental role in microtubule polymerisation. EB1 regulates the microtubule dynamic behaviour, through protein recruitment, and has been associated with several disease states, such as cancer and neuronal diseases. Diverse EB1 binding partners are recognised through a conserved SxIP motif within an intrinsically disordered region enriched with basic, serine and proline residues. Crystal structure of EB1 in complex with a peptide containing the SxIP motif demonstrated that the isoleucine-proline dipeptide is bound into a well-defined cavity of EB1 that may be suitable for small molecule targeting.

The research described herein reports the use of a multidisciplinary approach for the discovery of the first small molecule scaffold to target the EB1 recruiting domain. This approach included virtual screening (structure and ligand based design) and multiparameter compound selection. Solution NMR structures of the C-terminal domain of EB1 in the free form and in complex with the small molecule are also reported. A key finding from these structures is that the hydrophobic binding pocket reported to be essential for recruiting SxIP proteins is not pre-formed but highly dynamic in solution. This brings new insights to the protein recruitment mechanism regulated by EB1 and for the identification of new small molecule inhibitors for the EB1-SxIP protein interactions.

The interaction of short length peptides containing the SxIP motif with EB1 was characterised through the use of solution NMR and ITC methods. The contributions for the binding of the SxIP motif and neighbouring residues to EB1 were quantified in terms of binding energy. A structural model shows that the binding pocket of EB1 is largely extended when in complex.

This research describes not only the first chemical scaffold that targets EB1, it details important structural features of the interaction of this protein with SxIP containing peptides. This structural information provides fundamental understanding of this interaction that can be exploited in the future to discover higher affinity ligands.



# TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>iii</b>
<b>ABBREVIATIONS</b> .....	<b>ix</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
<b>1.1 Role of EB1 in +TIP targeting</b>	<b>1</b>
1.1.1 Microtubules .....	1
1.1.2 Microtubule plus end binding proteins - +TIPs .....	3
1.1.3 Disease states associated with EB1 .....	11
1.1.4 Structural features of EB1 .....	12
1.1.5 Interaction between EB1 and SxIP motif proteins.....	16
<b>1.2 Protein-protein interactions: a biological challenge for drug targeting</b>	<b>21</b>
1.2.1 Protein-protein interactions.....	21
1.2.2 Disorder and dynamics in protein-protein interactions.....	30
1.2.3 Screening techniques in drug discovery applied to protein-protein interactions .....	31
<b>1.3 Research Aims</b>	<b>38</b>
<b>Chapter 2. Identification of Novel Small Molecule Inhibitors</b>	<b>41</b>
<b>2.1 Virtual screening methods applied to identify hit molecules to target EB1</b>	<b>41</b>
<b>2.2 Identification of small molecules based on SxIP motif</b>	<b>44</b>
<b>2.3 Ligand-based design - Pharmacophore search</b>	<b>45</b>
<b>2.4 Molecular Docking studies</b>	<b>47</b>
2.4.1 Molecular Docking method validation and general considerations.....	47
2.4.2 Docking of virtual library obtained from ZINCPharmer .....	50
<b>2.5 Screening and selection of hit molecules</b>	<b>51</b>
2.5.1 Consensus scoring.....	51
2.5.2 Ligand Efficiency.....	51
2.5.3 Docking descriptors calculation.....	51
2.5.4 Multi-parameter optimization .....	52
2.5.5 Top ranked 100 molecules – selection of candidate compounds to synthesis and biological assays.....	53
2.5.6 Alternative compounds to ZINC12677264.....	59
<b>2.6 NMR based screening for hit candidates based on SxIP motif</b>	<b>63</b>
<b>Chapter 3. Structural characterisation of EB1c using solution NMR</b>	<b>72</b>
<b>3.1 Design of EB1 fragments</b>	<b>72</b>
<b>3.2 Production of recombinant EB1c<math>\Delta</math>8 and EB1c<math>\Delta</math>16</b>	<b>72</b>
<b>3.3 Characterisation of EB1c by NMR Spectroscopy</b>	<b>76</b>
<b>3.4 Solution NMR structure of EB1c<math>\Delta</math>8</b>	<b>82</b>
<b>3.5 Solution NMR structure of the complex EB1c<math>\Delta</math>8-1a</b>	<b>91</b>
<b>3.6 Analysis of the implications of the side chain dynamics for Arg222 and Tyr247 and its influence on small molecule binding</b>	<b>100</b>
3.6.1 Cross-docking approach.....	100
3.6.2 Principal component analysis.....	102

<b>Chapter 4. Design and identification of inhibitors beyond SxIP motif</b>	<b>109</b>
<b>4.1 Rational design of a hybrid molecule combining the SxIP mimetic scaffold and a tri-peptide</b>	<b>109</b>
4.1.1 NMR based screening of the hybrid molecule – 2a .....	112
4.1.2 Understanding of the loss of specificity to the IP binding site in the hybrid molecule .....	114
<b>4.2 Identification of small molecules based on the hybrid molecule</b>	<b>116</b>
4.2.1 Virtual screening .....	116
4.2.2 Selection of candidate molecules .....	119
4.2.3 NMR based screening for hit candidates based on the hybrid molecule .....	126
<b>Chapter 5. Biophysical Characterisation of Physiological EB1 interactions involving MACF2</b>	<b>133</b>
<b>5.1 NMR screening of MACF2 peptides against EB1cΔ8</b>	<b>133</b>
5.1.1 Chemical shift changes and mapping .....	140
5.1.2 Determination of binding affinity by NMR spectroscopy .....	145
<b>5.2 NMR screening of MACF peptides against EB1cΔ16</b>	<b>150</b>
5.2.1 Chemical shift changes and mapping .....	153
5.2.2 Determination of binding affinity by NMR spectroscopy .....	157
<b>5.3 Isothermal titration calorimetry binding assays of 11MACF - EB1cΔ8 and 11MACF - EB1cΔ16</b>	<b>160</b>
<b>5.4 Structural elucidation of the complex EB1cΔ8-11MACF using solution NMR164</b>	
<b>5.5 Optimisation of the native interaction</b>	<b>171</b>
5.5.1 Molecular modelling .....	171
5.5.2 NMR screening of MACF mutated peptides against EB1cΔ8 .....	176
5.5.3 Isothermal titration calorimetry binding assays of MACF mutated peptides ..	182
<b>5.6 Elucidation of the binding contributions to EB1c and the role of the C-terminus</b>	<b>186</b>
<b>Chapter 6. Conclusions and future work</b>	<b>191</b>
<b>Chapter 7. Materials and Methods</b>	<b>197</b>
<b>7.1 Molecular Modelling</b>	<b>197</b>
7.1.1 Pharmacophore search .....	197
7.1.2 EB1 Docking Protocol .....	197
7.1.3 Methods used for screening and selection of target molecules .....	198
7.1.4 QSAR experimental procedures .....	200
<b>7.2 Recombinant EB1 - cloning, expression and purification</b>	<b>201</b>
7.2.1 Materials .....	201
7.2.2 Methods .....	205
<b>7.3 Organic synthesis methods</b>	<b>214</b>
7.3.1 Materials .....	214
7.3.2 Methods .....	214
<b>7.4 Protein-ligand interactions</b>	<b>219</b>
7.4.1 NMR Spectroscopy .....	219
7.4.2 Isothermal Titration Calorimetry (ITC) .....	221
7.4.3 Octet® RED96 .....	221
7.4.4 NMR based structure calculation .....	222



REFERENCES.....	257
APPENDIX.....	xii



## ABBREVIATIONS

ADME - Absorption, Distribution, Metabolism and Excretion  
ADR - Ambiguous Distance Restraints  
APC - Adenomatous Polyposis Coli  
ARIA - Ambiguous Restraints for Iterative Assignment  
ASP - Astex Statistical Potential  
Bcl-2 - B-cell lymphoma 2  
CcpNmr Analysis - Collaborative Computational Project for NMR  
CH - Calponin Homology  
CLASP - Cytoplasmic Linker Associated Protein  
CLIP115 - Cytoplasmic Linker Protein of 115 kDa  
CLIP170 - Cytoplasmic Linker Protein of 170 kDa  
CNS - Crystallography & NMR System  
COSY - homonuclear COrrrelation SpectroscopY  
cryo-EM - Cryo-electron microscopy  
DANGLE - Dihedral ANgles from Global Likelihood Estimates  
DCC - N,N'-Dicyclohexylcarbodiimide  
DCM - Dichloromethane  
DIC - N,N' Diisopropylcarbodiimide  
DIEA - N,N-Diisopropylethylamine  
DMF - Dimethylformamide  
DMSO-d6 - Hexadeuterodimethyl sulfoxide  
dsBMA - Distal Spinal Bulbar Muscular Atrophy  
DTT - Dithiothreitol  
EB1 - End-Binding Protein 1  
EBH - EB Homology  
EDCI - 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide  
EDTA - Ethylenediaminetetraacetic acid  
Eq. - molar equivalent  
ESI - Electrospray Ionisation  
E-site - Nucleotide Exchangeable Site  
EtOAc - Ethyl acetate  
Fmoc - Fluorenylmethyloxycarbonyl chloride  
GA - Genetic Algorithm  
GDP - Guanosine Diphosphate  
 $\Delta G$  - free Gibbs energy

GOLD - Genetic Optimisation for Ligand Docking  
GPCR - G Protein-Coupled Receptor  
GTP - Guanosine Triphosphate  
HDM2 - Human Protein Double Minute 2  
 $\Delta H$  - Enthalpy  
HOBt - Hydroxybenzotriazole  
HPLC - High Pressure Liquid Chromatography  
HSQC - Heteronuclear Single Quantum Coherence Spectroscopy  
HTS – High-Throughput Screening  
IL-2 - Cytokine Interleukine-2  
ITC - Isothermal Titration Calorimetry  
 $K_a$  - association constant  
 $K_d$  - dissociation constant  
LCMS - Liquid Chromatography-Mass Spectrometry  
LE - Ligand Efficiency  
MACF - Microtubule-Actin Crosslinking Factor  
MAPs - Microtubule Associated Proteins  
MCAK - Mitotic Centromere-Associated Kinesin  
MeOH - Methanol  
MT - Microtubule  
NMR - Nuclear Magnetic Resonance  
NOE - Nuclear Overhauser Effect  
NOESY - Nuclear Overhauser Spectroscopy  
OPPF - Oxford Protein Production Facility  
PAGE - Polyacrylamide Gel Electrophoresis  
PCR - Polymerase Chain Reaction  
PDB - Protein Data Bank  
PLP - Piecewise Linear Potential  
PPI - Protein-Protein Interaction  
QSAR - Quantitative Structure-Activity Relationship  
RF - Random Forest  
RMSD - Root-mean-square deviation of atomic positions  
ROC - Receiver Operating Characteristic  
RSCB - Research Collaboratory for Structural Bioinformatics  
SAR - Structure-Activity Relationship  
SA - Simulating Annealing  
SAXS - Small-angle X-ray scattering

SDS - Sodium Dodecyl Sulfate  
SOC - Super Optimal broth with Catabolite Repression  
STIM1 - Stromal Interaction Molecule 1  
SUMO - Small Ubiquitin-like Modifier  
SVM - Support Vector Machine  
TCEP - Tris(2-carboxyethyl)phosphine  
TFE - Trifluoroethanol  
+TIPs - Plus-End-Binding Proteins  
TLC - Thin Layer Chromatography  
TOCSY - TOtal Correlation SpectroscopY  
TΔS - Entropy  
v/v - volume/volume  
w/v - weight/volume



## Chapter 1. INTRODUCTION

### 1.1 Role of EB1 in +TIP targeting

#### 1.1.1 Microtubules

Cells undergo substantial changes in shape as they grow, divide and move, and the cytoskeleton is the principal molecular machinery responsible for these functions.<sup>1</sup> The cytoskeleton due to its molecular complexity can be divided into three different functional subsystems – actin filaments, microtubules (MTs) and intermediate filaments.<sup>1</sup> These cytoskeletal components interact dynamically to control fundamental processes including mitosis, cell division, intracellular transport, cell motility and the regulation of cell organization.<sup>2</sup> To understand these mechanisms represents a challenge and is the focus of intense research. Furthermore, cytoskeletal proteins are implicated in many human pathological conditions, including cancer, infection, cardiovascular, inflammatory and neurodegenerative diseases.<sup>2</sup>

Many of the cytoskeletal activities rely on MTs.<sup>2</sup> These cellular polymers are key cytoskeleton components of eukaryotic cells playing important roles in many cellular processes, including intracellular transport, cell motility, meiosis and mitosis.<sup>3,4</sup> Their functions include cell shape, cytoplasm organization by positioning organelles, cell transport of other cellular entities and cell polarity. In an interphase cell the MT cytoskeleton is also critical for cell motility and a key constituent of cilia and flagella. During cell division, the MT rearrange into a spindle structure that segregates chromosomes.<sup>5</sup>

These functions rely on a specific MT ability of fast polymerization dynamics, with network-wide turnover rates on the order of minutes - known as “dynamic instability”.<sup>1,6</sup> This “dynamic instability” comprises periods of persistent microtubule growth interrupted by occasional switching to rapid shrinkage – microtubule catastrophe – and then by switching back from shrinkage to growth – microtubule rescue.<sup>3,7</sup>

Structurally, MTs are made from subunits of  $\alpha$  and  $\beta$ -tubulin heterodimers which assemble, in a head to tail manner, into a polar protofilament.<sup>5</sup> The lateral and parallel

association of ~13 protofilaments forms a hollow, polar cylinder. These lateral contacts are homotypic ( $\alpha$  -  $\alpha$  and  $\beta$  -  $\beta$  contacts), except at a single site or “seam” (with  $\alpha$  -  $\beta$  and  $\beta$  -  $\alpha$  contacts) (Figure 1.1).<sup>8</sup> The polymerisation occurs upon the binding of guanosine triphosphate (GTP) to the nucleotide exchangeable site (E-site) in  $\beta$ -tubulin. The GTP bound at the N-site (non-exchangeable) in  $\alpha$ -tubulin is buried at the intradimer interface where it plays a structural role.<sup>8</sup> Only dimers bound to GTP can polymerise, but upon polymerisation the GTP is readily hydrolysed to guanosine diphosphate (GDP) – making these subunits very unstable.<sup>9,10</sup> The MT structure is stabilised by the GTP subunits at the growing end and a possible loss of GTP-tubulin, allows the GDP-tubulin to curve, spreading out and the microtubule shrinks.<sup>9,11,12</sup> – Figure 1.2.

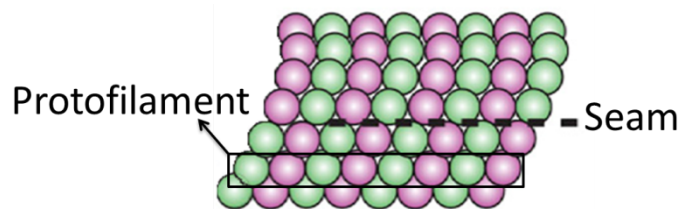


Figure 1.1 – Schematic representation of the MT structure.  $\beta$ -tubulin is represented in pink and  $\alpha$ -tubulin in green.

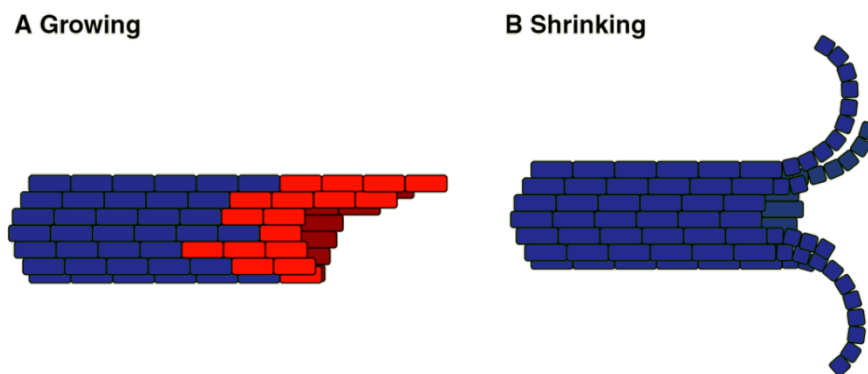


Figure 1.2 - Microtubules grow by addition of GTP-tubulin subunits (red) and microtubule shrinking by GDP-tubulin (blue) curving.

In most cells, MT polymerization is initiated at specific locations (*e.g.*, the centrosome), generating microtubule arrays of fixed polarity.<sup>13</sup> Polymerization is a polar process that reflects the polarity of the tubulin dimer, which in turn dictates the polarity of the MT. In other words, purified tubulin polymerizes more quickly from one end, the plus end, which is terminated by the  $\beta$ -subunit. The opposite end can slowly grow *in vitro*, though in cells it is usually stabilised or serves as the site of disassembly. It is known



as the minus end, and is ended by the  $\alpha$ -subunit.<sup>14</sup> Both MT ends can interact with sets of specific factors that control their dynamic status, intracellular localization and attachment to cellular structures.<sup>15</sup>

These cellular factors can stabilise or destabilise the MT polymer, controlling its dynamic behaviour and can be placed into two categories. The first category constituted by motors and microtubule associated proteins (MAPs) bind to MTs and alter MT properties in a non-covalent way; the second category includes tubulin post-translational modification enzymes that chemically modify the tubulin subunits.<sup>9,16</sup> For the scope of this project the focus will be only on the first category – MAPs.

Microtubule stability is controlled, to a large extent, by MAPs. For instance, the mitotic centromere-associated kinesins (MCAKs) bind to MT ends, destabilising them.<sup>5</sup> When MCAKs bind to MTs they induce the formation of the curl, weakening the association of the terminal GTP-tubulin dimer and promoting the release of GTP subunits from the end, which will release the GDP subunits constituent of the lattice.<sup>5,14</sup> On the other hand, the plus-end-binding proteins (or +TIPs), a structurally and functionally diverse group of proteins that are distinguished by their specific accumulation at microtubule plus ends, stabilise MTs during their growth phase.<sup>14,17</sup> +TIPs also regulate interactions between MTs and other intracellular structures, for instance the anchorage of microtubules the kinetochore of chromosomes, or organelles.<sup>12,18,19</sup> Loss or inhibition of these proteins compromises microtubule growth in many cell types.<sup>13</sup>

### 1.1.2 Microtubule plus end binding proteins - +TIPs

As just described, +TIPs are a diverse group of proteins that accumulate specifically at the MT plus ends (growing ends). There are two general mechanisms by which proteins can target the plus ends of MTs, vectorial transport and direct recruitment. They are not mutually exclusive, and both may contribute to the plus-end targeting of proteins.<sup>20</sup> First, kinesin motor proteins can power the vectorial transport of proteins along the MT body to the plus end. Second, a subset of +TIPs bind to unique structural features present at the plus end *versus* the body of the MT, and can recruit additional proteins to these sites. The prototypical example of this class of +TIPs are end-binding proteins (EBs), which are involved in recruiting binding partners.<sup>20</sup>

In order for +TIPs discriminate between MT lattice and the growing ends there must exist a structural difference. The presence of a GTP cap was initially reported as a possible structural feature for plus end recognition. However, +TIPs form a comet-like accumulation within 0.5–2  $\mu\text{m}$  distance of the centre of the microtubule, a region larger than the supposed length of the GTP cap. Alternatively, +TIPs might recognize tubulin sheets or individual protofilaments due to their curvature, or might bind certain tubulin sites, that are hidden inside the tube or obscured by inter-protofilament contacts in the rest of the microtubule.<sup>2</sup>

Since the discovery of cytoplasmic linker protein of 170 kDa - CLIP170, the first identified +TIP, several structurally unrelated families of +TIPs have been identified.<sup>2,14</sup> There are more than 20 families of +TIPs, ranging in size from a few hundred up to thousands of residues, multi-domain and/or multi-subunit. Despite their diversity, they often co-localise and share common activities, being difficult to classify in terms of function. Alternatively, conserved modular binding domains, repeat sequences and motifs can be used to group these proteins.

Table 1.1 - Principal +TIPs families and main interactions.<sup>12</sup>

Family	Examples	Interactions
EB proteins	EB1 EB2 EB3	Other +TIPs
CAP-Gly domain proteins	CLIP-170 CLIP-115 p150glued KIF13B	EB1, CLIP-170, CLIP-115, p150glued, MCAK, CLASP1 and 2 EB1, CLIP-170, CLASP1 and 2 EB1, CLIP-170
SxIP proteins	CLASP1/CLASP2 APC MACF STIM1 MCAK	EB1, CLIP-170, CLIP-115, ACF7 EB1, MCAK EB1, CLASP1 and 2 EB1 EB1, CLIP-170, APC
TOG domain	CLASPs XMAP215/Dis1	EB1

The detailed nature, specificity, and modes of regulation of their interactions are largely unknown. It has recently been established, however, that the protein-protein interaction scaffold of +TIP networks relies on a small number of protein modules and linear sequence motifs that are highly conserved throughout eukaryotic organisms. Prominent examples are the EB-like domain of end binding proteins, the cytoskeleton-associated protein glycine-rich (CAP-Gly) domain, and short and acidic carboxy-terminal sequence motifs – SxIP proteins. Therefore, the simplest way to classify

+TIPs into families is by using their domain composition – Table 1.1.<sup>2,18,21</sup>

Figure 1.3 describes the main domains and +TIPs that will be described in the next sub-sections.

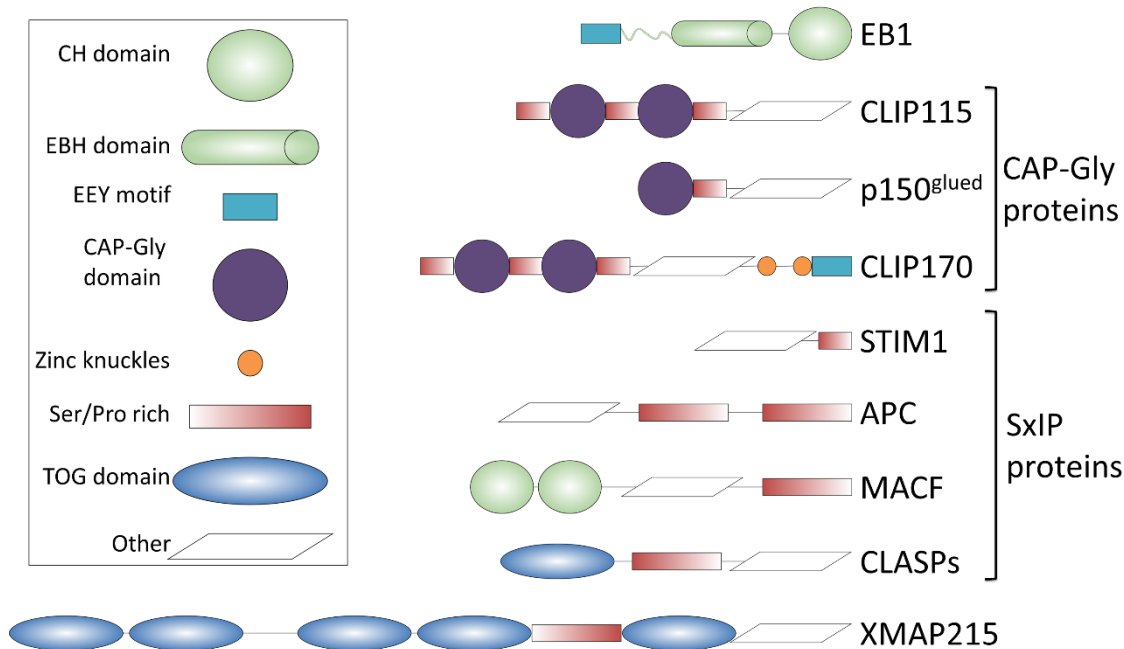


Figure 1.3 – Main +TIPs and their conserved domains. White polygon means any other domains/structural features that are relevant for +TIPs interactions networks.

### 1.1.2.1 End binding proteins

Three end-binding family proteins – EB1, EB2 and EB3, have been reported. EB family members contain an N-terminal calponin homology (CH) domain that directly recognizes a structural feature of growing microtubule ends, and a C-terminal EB homology (EBH) domain that mediates binding to other +TIPs.<sup>2,22</sup> EB1 was shown to promote MT polymerization and interacts directly with many other +TIPs and cytoskeletal proteins such as CLIP170, the dynactin large subunit p150<sup>Glued</sup>, the mitotic centromere-associated kinesin (MCAK), the microtubule-actin crosslinking factor (MACF) and the adenomatous polyposis coli (APC).<sup>23</sup> EB1 and EB3 share a high sequence conservation and can form heterodimers.<sup>24</sup> Moreover, EB3 binding site is very similar to the one observed for EB1 as demonstrated by the structure published in 2012 by Bjelić and co-workers – PDB code 3TQ7.<sup>24</sup> EB2, opposite to what was observed to EB1 and EB3, does not bind to APC, indicating possible structural differences.

In terms of expression in different cell lines there are also differences. EB1 is expressed at constant levels in a variety of cell lines, whereas EB2 and EB3 vary. EB3 is especially abundant in the central nervous system and in muscles.<sup>17,25</sup> It is not completely clear what are the roles of EB2 and EB3, but all three end-binding proteins are expressed in mammalian cells. Depletion of EB1 had no effect on EB3 localisation, but led to the distribution of EB2. Depletion both EB2 and EB3 had no effect on the distribution of the remaining EBs. Finally, simultaneous depletion of EB1 and EB3 caused a redistribution of EB2 similar to the one induced by depletion of EB1 alone.<sup>26</sup>

### 1.1.2.2 CAP-Gly domain proteins

The cytoskeleton-associated protein glycine rich (CAP-Gly) domain-containing proteins have an N-terminal CAP-Gly globular domain containing a unique hydrophobic cavity that encompasses the highly conserved GKNDG sequence motif and several characteristic glycine residues. The C-terminal region, can recruit various cargos, namely for CLIP170 that contains two tandemly repeated metal binding motifs (zinc knuckles) and a C-terminal EEY/F. The EEY/F motif is present in the end of the C-terminus tail of EB1 and was shown to be a recruiting domain for CAP-Gly domains.<sup>2,21,27</sup>

Mutations in the CAP-Gly domain of p150<sup>Glued</sup> subunit of dynactin are associated with various neurological disorders. For example, a G59S point mutation has been identified in patients with distal spinal bulbar muscular atrophy (dSBMA). *In vivo* and *in vitro* studies suggest that this mutation causes misfolding of the CAP-Gly motif, and reduces the binding affinity of the mutant to microtubules. Five other mutations, G71A, G71E, G71R, T72P and Q74P, were identified in patients diagnosed with Perry's syndrome, a severe neurological disease whose manifestations are parkinsonism and weight loss accompanied by depression, social withdrawal and suicidal attempts. It was firstly hypothesized that, as the result of all these mutations, binding to microtubules was disrupted, but more recent studies indicate that the binding affinity to microtubules is very similar or even higher for these mutants. Instead, the binding to EB1 is abrogated.<sup>28</sup>

The CAP-Gly domain seems to have two distinct regions that bind to EB1. The first comprises the GKNDG motif and several aromatic residues, Phe252, Trp57 and Phe88. This site, that has been predicted to interact with MTs, is also shown to be in contact

with the EEY motif of EB1. The second interface consists of the  $\beta$ 2- $\beta$ 3 loop, <sup>49</sup>ATLFLAT<sup>54</sup> showing interactions with the EB1-like motif, *i.e.* the hydrophobic cavity of EB1, with the Ala49 inserted in the pocket.

Despite the existence of a binding site to MTs in the CLAP-Gly domain, the complete deletion of the binding sites in EB1 results in activation of MT polymerisation by EB1 alone. An explanation is possibly the fact that the EB1 C-terminal acidic tail has a unique sequence – <sup>262</sup>QEEQEEY<sup>268</sup> – very similar to the  $\alpha$ -tubulin C-terminus – <sup>445</sup>EEEGEEY<sup>451</sup>, and if the region that recognises this sequence of EB1 is deleted the MT recognition region is also lost.<sup>29</sup>

### 1.1.2.3 SxIP proteins

Several +TIPs do not appear to use a defined and structured domain for targeting the MT ends. Instead, they contain extensive sequence regions that are enriched with basic and serine residues. These regions, which are predicted to be flexible, contain a EB1-binding motif – SxIP, and track the plus ends with EB1 dependency.<sup>12,30,31</sup> These are known as the SxIP proteins. Included in this vast group one can find the cytoplasmic linker associated proteins (CLASPs), APC, MACF, the calcium sensor (STIM1) and MCAK.

Sequence analysis for a large number of these proteins revealed that they are very diverse in terms of domain composition but contain low complexity sequence regions, rich in basic, serine and proline residues, usually towards the C-terminus. The SxIP motif also termed as *microtubule tip localization signal*, also shows some variability across SxIP proteins<sup>21,31,32</sup> and it is likely to occur in many more proteins than the +TIPs identified so far.<sup>12</sup> The isoleucine (or leucine) and proline residues are involved in hydrophobic interactions and are the most important, being also more conserved.<sup>22</sup> The serine at position 1 contributes to hydrogen bonds with conserved residues surrounding the EB hydrophobic groove, but recent data indicate that it is not required for plus-end tracking.<sup>22</sup> Table 3 demonstrates this variability in known, probable and potential SxIP motifs found in a wide range of proteins with different cellular functions.

Less well conserved variations of the SxIP motif can mediate plus-end tracking, and many +TIPs contain multiple functional SxIP motifs that act in concert and increase the affinity for EB1. CLASP2, for example, contains two adjacent SxIP motifs, which increases the affinity of this +TIP for growing microtubule plus ends. CLASP2 is

monomeric and one CLASP2 molecule is therefore able to bind to the two EBH domains in an EB dimer.<sup>12,26</sup>

Table 1.2 – EB1 recruited +TIPs containing confirmed or potential SxIP motifs. Adapted from Kumar & Wittmann, 2012.<sup>22</sup>

Protein	SxIP motifs	Uniprot identifier	Proposed functions
<b>+TIPs with confirmed SxIP motifs</b>			
CLASP1, CLASP2	SKIP SRIP	Q7Z460 O75122	MT capture, dynamics and nucleation at the Golgi apparatus
MACF1	SKIP	Q9UPN3	MT capture
APC	SQIP	P25054	MT capture; mutated in colon adenocarcinoma
STIM1	TRIP	Q13586	ER-microtubule interactions; activator of store-operated Ca <sup>2+</sup> entry; mutated in particular immunodeficiencies
MCAK	SKIP	Q996610	MT depolymerase
KIF18B	SFLP SSLP GGIP SAIP	Q86Y91	MT depolymerase
SLAIN2	SGLP GGIP RSLP	Q9P270	MT dynamics; recruits XMAP215/ch-TOG to MT plus ends
Sentin/SSP2 (Drosophila)	TGIP	QVUA5	MT dynamics; recruits XMAP215/ch-TOG homolog to MT plus ends
CDK5RAP2	SRLP	Q96SN8	Centrosome maturation; mutated in autosomal recessive primary microcephaly
Kebab (Drosophila)	TKIP	Q9VQ69	Kinetocore component; unknown function
Ipl1 (S. cerevisiae)	SKIP SKIP	P38991	Yeast Aurora kinase homolog
<b>+TIPs with probable SxIP motif</b>			
Melanophilin	SNLP	Q9BV36	Melanosome transport; mutated in Griscelli syndrome
P140CAP	TSIP	Q9C0H9	MT interactions in dendritic spines; Src kinase regulator
MTUS2/TIP	SRLP SNLP SRLP SLLP	Q5JR59	MT dynamics; recruits MCAK to MT plus-ends
FOP	SKIP	Q95684	Centrosome MT anchoring; mutated in some myeloid cancers
NAVIGATOR1	SRIP SGIP SLIP	Q8NEY1 (Q8IVL1, Q8IVL0)	Neurite outgrowth
DRhoGEF2	SKIP	Q44381	RhoGTPase signaling
<b>Potential SxIP motif-containing EB-binding proteins; unconfirmed plus-end tracking activity</b>			
PSRC1/DDA3	SAIP	Q6PGN9	Mitotic regulator of MCAK
Plexin-B3	SGIP	Q9ULL4	Axon guidance

Honnapa and co-workers<sup>31</sup>, in 2009, assessed the role of the two SxIP motifs by replacing the Ile-Pro dipeptide of the two SxIP motifs individually or simultaneously by Ser-Ser. The replacement into either of the SxIP motif alone did not reduce binding to EB1, but significantly reduced accumulation of CLASP2 mutants at MT plus ends. When both Ile-Pro were mutated, CLASP2 had a strong reduction in EB1 binding and failed to localize to growing MT ends.<sup>31</sup> These findings show that the presence of two SxIP motifs increases the affinity of +TIPs for the growing MT ends.

The mode of interaction of these proteins with EB1 will be more detailed in section 1.1.5, as it will be one of the main focus of this project.

#### ***1.1.2.4 TOG domain proteins***

In addition to the SxIP motif, CLASPs have a Tumour-overexpressed gene (TOG) like domain, and are able to bind to tubulin through these domains. TOG domains are 250 residue repeats found in several MAPs families allowing direct binding to MTs contrarily to what happens to SxIP proteins.

Members of the XMAP215/DIS1 family contain several TOG domains in their N-terminal ends. XMAP215, which contains five TOG domains at its N terminus, functions as a processive MT polymerase, binding to tubulin dimers, through a conserved flat tubulin flat surface, to facilitate their incorporation into MT plus-ends.<sup>33,34</sup>

TOG-like domains in CLASPs have a convex tubulin-binding surface, suggesting that the recognition mechanism of tubulin by CLASPs differs from that of the XMAP215/Dis family. Functions of CLASPs are also different from XMAP215/Dis1 families, including MT rescue and suppressing MT catastrophe events.<sup>33,35</sup>

#### ***1.1.2.5 Main interactions at microtubule plus ends mediated by +TIPS***

As described in the previous sections, EB1 acts as a master regulator at MT plus ends protein interactions. However, CAP-Gly proteins despite having two binding motifs that bind to EB1 – GKNDG and ATLFLAT - can autonomously target MTs through the recognition of the EEY motif in tubulin - Figure 1.4. Both EBs and CLIPs are dimers and seem to require at least two tubulin-binding domains (CH or CAP-Gly, respectively) to track microtubule ends, suggesting that the affinity of individual sites

for microtubules or tubulin is relatively weak. The arrangement of these sites with respect to each other does not seem to be important because an artificial combination of a CH domain and a CAP-Gly domain is capable of plus-end tracking.<sup>2</sup> CLIP170 has the EEY/F motif, meaning that, similarly to EB1, it can recruit other CAP-Gly domains. Additionally, CLIP170's zinc knuckles can bind to CLIP170 and p150.

CLASPs due to their TOG-like domain can directly bind to MTs. They can also bind to EB1 using their SxIP motif. In CLASP-depleted cells, EBs localize along the MT lattice in addition to plus ends. The MT-binding region of CLASP was sufficient for restoring normal EB localization.<sup>36</sup> Similarly, XMAP215/DIS1 uses its TOG domain to bind directly to MTs.

Finally, most of SxIP protein can only track the MTs plus ends through EB1 binding. These interactions are summarised in Figure 1.4 and Table 1.3.

Table 1.3 – Summary of known domains and motifs interacting within the +TIP network. Adapted from Duellberg et al., 2014.<sup>29</sup>

Domain/motif	Found in	Interacts with
CH domain	N-terminal part of EBs	Growing MT plus ends
EBH domain	Part of C-terminal region of EBs	SxIP motifs, CAP-Gly domains of p150
EEY/F motif	C-terminus of EBs, CLIP170 and $\alpha$ -tubulin CLIP170	CAP-Gly domains
Zinc knuckle	CLIP170	CAP-Gly domains of p150 and CLIP-170
SxIP motif	Several unrelated EB-dependent +TIPs	EBH domains
CAP-Gly domain	CLIP170, CLIP115 and p150	EEY/F motifs and EBH domains



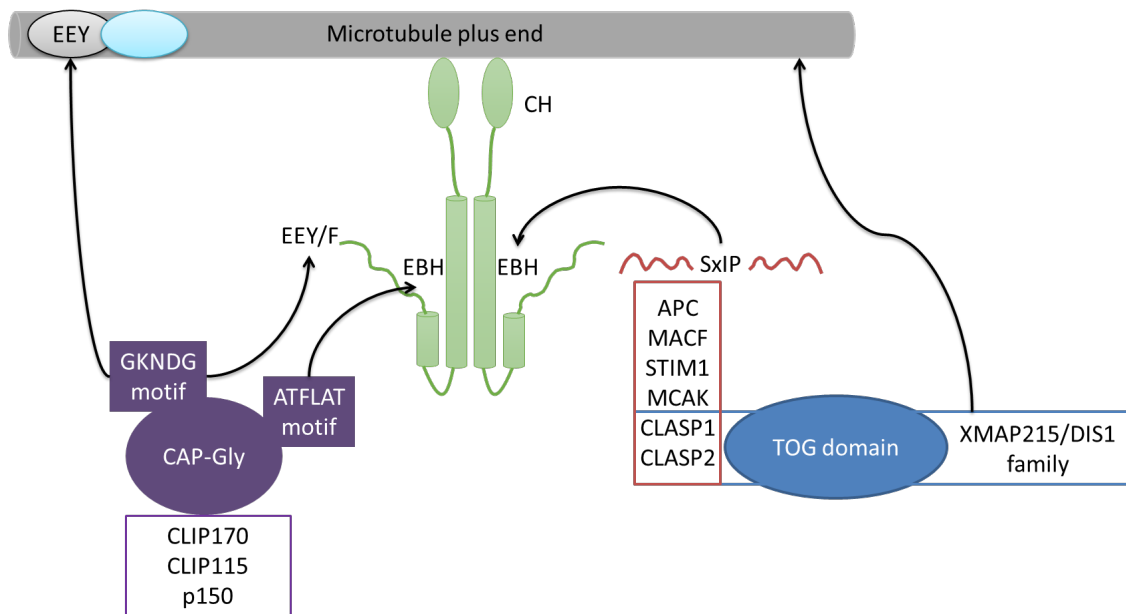


Figure 1.4 – Schematic representation of possible +TIPs networks at the MT plus ends. Binding of SxIP proteins to EBH domain is one of the most well documented interaction, followed by the CAP-Gly domain proteins interactions with the EEY motif of EB1. The interaction of the ATFLAT motif from CAP-Gly proteins with the EBH domain of EB1 has also been reported. EB1 is shown in green, SxIP proteins in red, TOG domain in blue, CAP-Gly proteins in purple.  $\alpha$ -tubulin dimer is showed in grey and  $\beta$ -tubulin in light blue.

Due to the importance of EB1 as a recruiting protein to the MT plus ends, a more detailed description of the structure features will be subsequently described.

### 1.1.3 Disease states associated with EB1

EB1 was initially described as a protein that interacts with the C-terminus of APC that has a known tumour suppressor function and when mutated may result in colorectal cancer.<sup>12,37</sup> Observations revealed that these mutations resulted, very often, in the truncation of APC leading to the loss of the region that interacts with EB1. It was, therefore, postulated that EB1-APC interaction may be important for the tumour suppression activity of APC.<sup>38</sup> Later, EB1 overexpression has been reported in gastric adenocarcinoma, hepatocellular carcinoma, oesophageal squamous cell carcinoma and breast cancers.<sup>37</sup> Conversely, the knockdown of EB1 expression promoted an increase in apoptosis. These functions were associated with the activation of the  $\beta$ -catenin pathway – promoting the growth and inhibiting apoptosis.<sup>25,37,39</sup>

In addition to the above-mentioned relationship between EB1, APC and cancer/apoptosis regulation, EB1 deregulation can influence other biological processes

as it recruits a wide number of partners to the MT plus-ends.<sup>40</sup>

Another relevant example is MACF/ACF7, member of the spectraplakin family that through their binding to EB1 reinforce links between F-actin and MTs, acting as an integrator of MT-actin dynamics.<sup>41</sup> Ablation of the ACF7 gene in mice results in embryonic lethality and various medical conditions and development defect arise from mutations in genes encoding spectraplakins, such as mental retardation, cancer and chronic skin blistering.<sup>41</sup>

#### 1.1.4 Structural features of EB1

EB1 binds to the MT plus ends with higher affinity (about 10-fold higher) when compared to the MT lattice, presenting also fast binding/unbinding kinetics.<sup>19</sup> These characteristics are fundamental for the EB1's functions at the MT plus ends. This protein has been structurally characterised by X-ray crystallography (both N and C-terminal domains), nuclear magnetic resonance (NMR) spectroscopy (N-terminal domain) and small angle X-ray scattering (SAXS), combined with single-particle electron microscopy (for the model of the entire EB1 molecule).<sup>22,42</sup> SAXS data shows the overall structure consists of two large domains, N- and C-terminal, joined via a less conserved linker sequence, forming a flexible Y shaped structure (Figure 1.5).<sup>2,42,43</sup> The same low resolution model of the EB1 molecule shows that despite being a homodimer, EB1 can be surprisingly asymmetric. The C terminal rod-like coiled coil is more closely associated with one of the CH domains and protrudes from the microtubule-binding domains at an angle of almost 45°.<sup>22</sup>

The N-terminal domain is necessary and sufficient for MT binding. Its structure, solved by X-ray crystallography (PDB codes 2R8U (Huang, Lovelace, Smith, & Lebioda, unpublished), 1PA7 and 1UEG<sup>45</sup> and solution NMR (PDB code 1V5K (Tomizawa, Kigawa, Koshiba, Inoue, & Yokoyama, unpublished), revealed a CH fold and consists of six helices.<sup>23,43</sup> The architecture of the domain is dominated by four major  $\alpha$ -helices ( $\alpha 1$ ,  $\alpha 3$ ,  $\alpha 4$  and  $\alpha 6$ ). The first helix ( $\alpha 1$ ) forms an angle of  $\sim 75^\circ$  with the central helices  $\alpha 3$  and  $\alpha 4$ . Three helices  $\alpha 3$ ,  $\alpha 4$ , and  $\alpha 6$  form a parallel three helices bundle, giving rise to a hydrophobic core,  $\alpha 4$  and  $\alpha 6$  are partially exposed to the solvent, creating a conserved hydrophobic cleft that provides a potential protein-protein interaction surface.

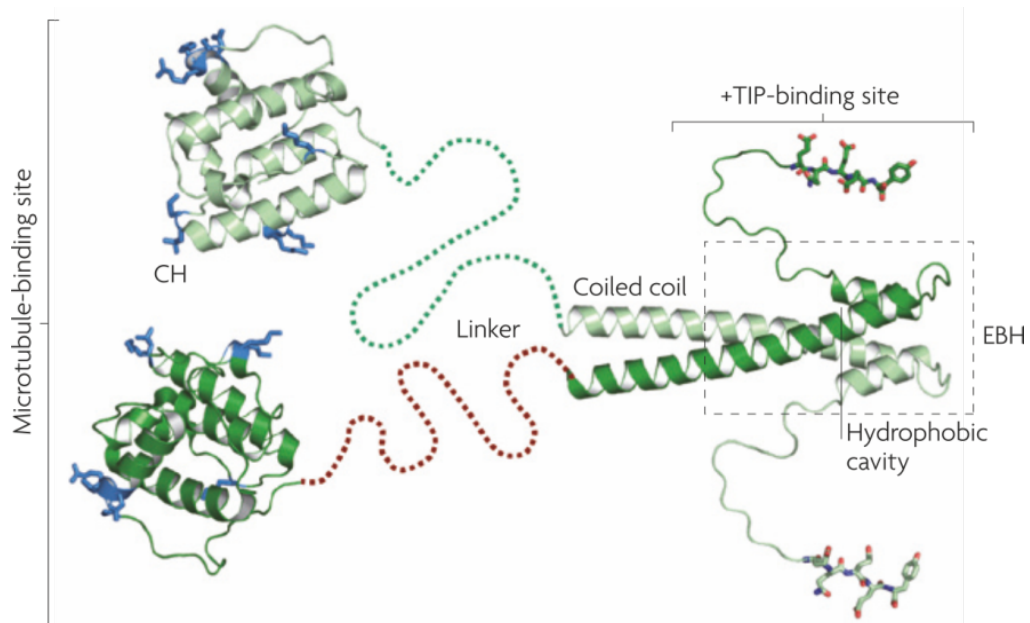


Figure 1.5 - Representation of the overall structure of EB1, composed by two larger globular domains (N-terminal), joined via a flexible linker to a coiled-coil domain (C-terminal). This figure is a reproduction of the one published by Akhmanova & Steinmetz, 2008.<sup>2</sup>

EB1 (and its fission yeast homolog Mal3) targets to growing MT ends by recognizing a nucleotide-dependent structural state. The binding mode of N-terminal domain of Mal3, the fission yeast EB, to MT plus ends was described as the interaction with four different tubulin dimers – Figure 1.6.

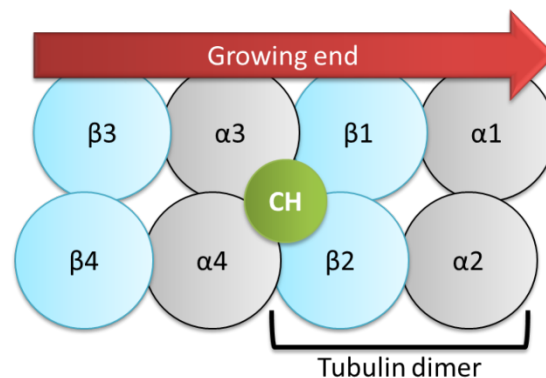


Figure 1.6 – Suggested binding mode of EB1 CH domain to microtubules. The binding site is formed by two adjacent  $\alpha$ -tubulin contacts (towards the microtubule plus end) and two adjacent  $\beta$ -tubulin contacts (towards the minus end).

The CH domain binding site is formed by two adjacent  $\alpha$ -tubulin contacts (towards the microtubule plus end) and two adjacent  $\beta$ -tubulin contacts (towards the minus end) showing that Mal3 binds to four neighbouring tubulins, at the junction between two protofilaments and two longitudinal interfaces.<sup>19</sup> Once polymerization occurs, the

interaction partners  $\alpha$  and  $\beta$  tubulin align themselves in a favourable position and EB1 may bind to these, stabilising the MTs promoting its MT polymerization.<sup>23</sup> Cryo-EM structures of MTs co-assembled with EB3 show that the presence of EB promotes a compacted MT lattice with a unique lattice twist, which EB itself preferentially binds. It was proposed that stabilisation of MTs catalysed by EB is due to the strengthening lateral and longitudinal interactions between dimers that promote a more compact MT lattice. The compacted lattice could then speed up GTP hydrolysis. The strain generated in the tubulin subunit following GTP hydrolysis, will promote the breakage of lateral contacts, likely initiated at the weaker seam interface. EBs promote MT growth by binding at the intersection of lateral and longitudinal contacts and enforcing the proper geometry of contacts to facilitate seam closure.<sup>8</sup>

The C-terminal domain of EB1 (EB1c) contains a parallel two-stranded coiled coil followed by the unique EB homology domain (EBH) domain containing a four-helix bundle and a disordered C-terminal region encompassing the EEY/F sequence motif (Figure 1.7).<sup>31</sup>

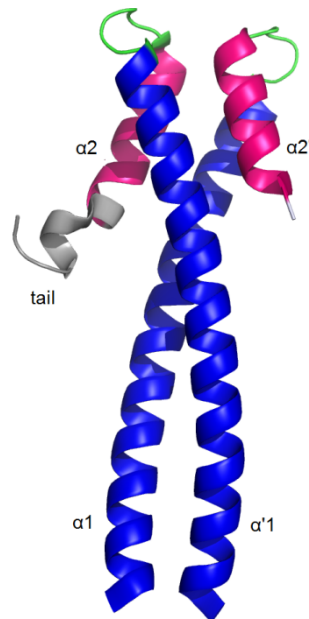


Figure 1.7 - Four-helix bundled structure of the C-terminal of EB1 (PDB code 1WU9). Coloured in blue are the long helices ( $\alpha 1$  and  $\alpha 1'$ ), followed by a hairpin connection (coloured green) leading to a short second helix ( $\alpha 2$  and  $\alpha 2'$ , coloured pink). Finally the C-terminal disordered tail is coloured grey.<sup>43</sup>

The two parallel  $\alpha 1$ -helices of EB1c wrap around each other in a slightly left-handed supercoil. The residues occupying the core *a* and *d* positions of the four heptad repeats (Figure 1.8) between Ala193 and Leu221 pack in a typical “knobs-into-hole” fashion.



Figure 1.8 – Sequence for the C-terminal domain of EB1 (EB1c). Coiled coil region shown in blue with positions *a* and *d* of the heptad repeats underlined. The connection hairpin between the two helices is represented in green. The shorter helix is shown in pink followed by the disordered C-terminus in grey. Uniprot<sup>47</sup> code MAPRE1.

The EBH domain is responsible for the dimerization of EB1 monomers and represents the +TIPs recruitment domain of EB1. This domain has a four-helix bundled structure, where the two coiled-coil helices, residues 219-229 ( $\alpha$ 1), diverge into a fork like structure. The following C-terminal helices, residues 237-247 ( $\alpha$ 2), of each monomer form a complementary fork, joining with the  $\alpha$ 1 helices (Figure 1.7).<sup>23</sup> The large hydrophobic surface buried in this bundle is expected to significantly contribute to the stability of the dimeric structure of EB1c.<sup>43</sup> First, the side chains forming the hydrophobic core of this bundle (equivalent residue pairs are 221/239, 224/242 and 227/245) are highly conserved, additionally, mutation of Ile224 to alanine was shown to destabilise the four helical bundle and consequently abolish dimerization.<sup>27</sup>

Residues 211-229 of the N-terminal region of the coiled coil are highly conserved across 13 representative EB1 family members, from yeast to vertebrates. This region includes three invariant residues Phe216, Tyr217 and Phe218', forming the FYF' motif. The configuration of the coiled coil dictates that the spatial FYF' motif is comprised of Phe216 and Tyr217 from one chain and Phe218' from the other chain of the dimer. Additional conserved residues at this junction are two solvent exposed residues of the  $\alpha$ 2 helix: Ile245 and Tyr247. This conserved surface is characterized by a hydrophobic character and a negative net charge contributed by the acidic residues Glu211, Glu213, Asp215 and Glu225. To evaluate the individual contribution of residues within the 211-229 conserved region, single alanine mutations were prepared and mutant EB1 proteins were analysed for MACF2<sub>1595-1637</sub> (SxIP containing protein) binding. The analysis revealed Glu211, Asp215, Phe216, Tyr217, Phe218, Arg222, Glu225 and Leu226 as the key binding residues. The FYF' motif provided a critical contribution to the binding interaction since mutation of any of these three residues abolish binding - Figure 1.9.<sup>27</sup>

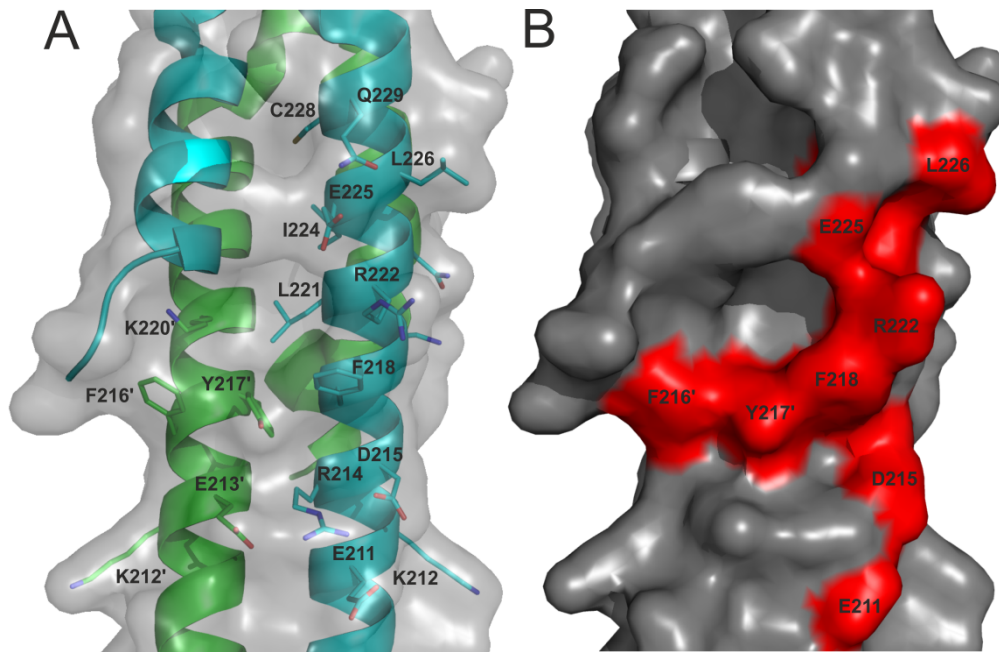


Figure 1.9 - Surface and cartoon representation of EB1. A - conserved residues are explicitly shown as sticks for a better visualisation and each monomer coloured differently, cyan and green.<sup>48</sup> B – surface representation where the residues coloured in red were shown to be important for the binding to MACF21595-1637, through single alanine mutagenesis experiments, as reported by Slep and co-workers.<sup>27</sup>

Finally, the flexible C-terminal (Tyr247-Tyr268) tail of EB1 is unstructured and flexible, having a highly conserved acidic-aromatic C-terminus – the EEY/F sequence motif that is reminiscent of motifs that are found in  $\alpha$ -tubulin and CLIP170.<sup>2</sup> The entirety of the C-terminal disordered region was never characterised since the EBH domain of EB1 has only been studied by X-ray crystallography, unsuitable for unstructured regions.

### 1.1.5 Interaction between EB1 and SxIP motif proteins

MT plus-ends protein interactions are very complex and not fully understood as previously described. SxIP proteins track the MT plus ends through EB1 recruitment and relevant examples of this interaction are MACF, STIM1, APC and CLASPs.<sup>5</sup>

Crystal structure of EB1c $\Delta$ 8-MACFp1 complex (PDB code 3GJO<sup>31</sup>) provides details on the SxIP motif recognition. In this crystal structure EB1c domain (residues 191-268) and a short peptide (30 amino acid residues), derived from human MACF2, were used. There are two EBH homodimers in the asymmetric unit and consequently four independent binding sites. MACF2 binding region to EB1 is known as unstructured,

and from the 30 amino acids a maximum of 11 can be observed in one of the binding sites. Residues outside this SxIP-containing region are less well defined and do not participate in specific intermolecular interactions.<sup>31</sup> A heptapeptide <sup>5476</sup>PSKIPTP<sup>5482</sup> of the ligand is in a very similar conformation for three of them, meaning this sequence becomes more structured and less dynamic when bound to EB1 and therefore must be of great importance for the interaction. Residues of the SxIP motif make specific contacts with the protein. Ser5477 forms an extensive network of hydrogen bonds with the highly conserved Arg222, Glu225, Gln229 and Tyr247 residues of EB1 and a water molecule. The dipeptide Ile-Pro is buried within a cleft shaped by the residues Phe216-Arg222, Glu225, Leu246 and Tyr247 of the hydrophobic cavity and Ala248-Pro256 of the C-terminal tail. Further contacts between Thr5481 and Pro5482 with Phe253, Gly252 and Val254 are shown (Figure 1.10).<sup>31</sup>

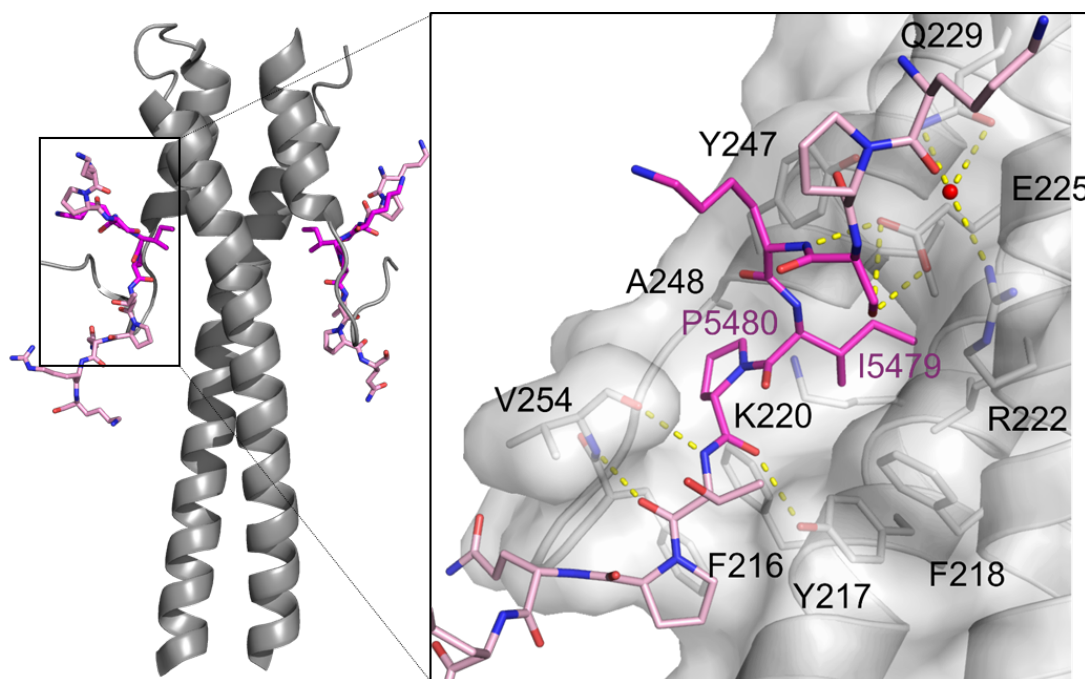


Figure 1.10 - Representation of the SxIP binding mode as shown by Honnappa *et al.*<sup>31</sup>. SxIP containing peptide is shown as light pink sticks, and the SxIP motif is highlighted as bright pink sticks. EB1 is overall shown as cartoon and surface representation with important residues shown as sticks. The sign ' is used after the residue name to indicate it belongs to the other monomer.

Buey *et al.* in 2012 made a systematic analysis of the important residues for the binding based on a 30 residue peptide derived from MACF2 – MACFp1 (the same fragment used to obtain the complex published by Honnappa *et al.*, 2009<sup>49</sup>) by mutating each one of the 30 positions at a time. They found that the 12 residues spanning the sequence region <sup>5476</sup>PSKIPTPQRKSP<sup>5487</sup> have a major contribution to the binding to EBH

domain of EB1. This region obviously includes the SxIP motif (SKIP), although it is noteworthy that the replacement of residues in this sequence by negatively charged glutamates or aspartates severely impaired the binding. Basic lysine and arginine residues are preferred, inclusively for the x position in the SxIP. Notably, this position does not tolerate many mutations, being arginine the only substitution that does not interfere with the binding. The side chain of this lysine is not pointing directly to the EB1 surface; instead, it seems to promote long range electrostatic attractive interactions with the negatively charged surface of the flexible C-terminal of EB1, e.g. Glu258. From the analysis of the possible replacements in the first position of SxIP, serine can only be replaced by threonine that also includes a hydroxyl group in its side chain. Regarding isoleucine at position 3, it only bears the replacement by the hydrophobic side chain of leucine, with significant reduction of the binding affinity. Finally, proline in position 4 cannot be replaced by any amino acid. The ring fills a hydrophobic pocket, and any other side chain would have the polar amide group from the backbone pointing into this hydrophobic region.<sup>49</sup> These observations are supported by a sequence alignment of SxIP proteins in Figure 1.11. Sequences surrounding functional SxIP motifs are interspersed with positively charged amino acids and are particularly enriched in arginine residues. Sequence alignment of 30 amino acid stretches around the reported *microtubule tip localization signal* (SxIP) shows the prevalence of serine and proline residues around these regions. There is also a very low content in negatively charged residues, with higher occurrence of positively charged residues (coloured in blue).

SxIP motifs are required for plus-end tracking, since the mutation of this motif abolishes binding to EB1, but by themselves are not sufficient. The hydrophobic interaction surface of the SxIP motif, and consequently the binding free energy contribution of this interaction, are probably fairly small.<sup>22</sup>





Figure 1.11 - Sequence alignment for known SxIP proteins based on a 30 residue sequence encompassing the identified SxIP motif. The sequence is colour coded as: positively charged residues (arginine and lysine) coloured in blue, serine and threonine coloured in green, hydrophobic residues (isoleucine and leucine) coloured in yellow and proline coloured in orange. Intensity of the colour indicates the higher conservation in that position. Figure was made using JalView 2.8.2.

The structure of EB1 in complex with the peptide derived from MACF2 also demonstrates an additional feature of the interaction. Residues 250-257, which are not resolved in the crystal structure of the free EB1c dimer, become structured in complex and form a handle-shaped loop wrapped around the bound peptide.<sup>31</sup> The EB1 C-terminus region (Thr249-Tyr268) contains negatively charged residues, in a 19 residue span, eight are negatively charged (Figure 1.8). SxIP proteins have high propensity to have positively charged residues around the SxIP motif (Figure 1.11) and it is reasonable to assume that electrostatic interactions contribute significantly to +TIP interactions with EBs.<sup>22</sup> In the case of CLASP2, there are indications that conserved arginine residues surrounding the SxIP motifs participate in extensive salt-bridge networks with negatively charged glutamates near the EB1 C-terminus. Such arginine–glutamate salt bridges are highly favourable, but they can be efficiently disrupted by phosphorylation.<sup>22</sup> These electrostatic interactions can be tuned by multisite phosphorylation through induction of intermolecular phosphoserine-arginine salt-bridges as an alternative to the intermolecular interactions with EB1 glutamate residues, and this can be a way of controlling these dynamic networks.<sup>32</sup> Therefore, the recruitment of SxIP motif containing +TIPS through EB1 binding depends on hydrophobic interactions of the SxIP motif, that determine specificity and on less specific electrostatic interactions that contribute largely to the binding free energy.<sup>22</sup>

The SxIP containing regions are within intrinsically disordered regions (IDRs). GTSE1, for instance, a protein which secondary structure prediction revealed that it is a mostly intrinsically disordered protein (IDP), contains multiple potential “SKIP”-like (SxLP) EB1-interaction motifs surrounded by basic residues.<sup>50</sup> At the same time, the C-terminal region following the EBH domain of EB1 and involved in the interaction with SxIP ligands is highly flexible and disordered. It has been reported that complexes involving IDRs are often highly dynamic and short lived.<sup>51</sup> IDRs often participate in interactions with high specificity but low affinity ( $\mu\text{M}$  range), facilitating rapid exchange of binding sites between multiple interacting partners. These observations are in agreement with what has been reported for the SxIP-EB1 interactions, where the EB1 binding affinities of many SxIP containing regions are in the low/medium micromolar range.<sup>49</sup>

For the +TIPs networks mediated by EB1 the existence of IDRs seems to be fundamental. The inherent flexibility of IDRs allows for a degree of promiscuity in

interactions with cellular partners containing the SxIP motif that will give the specificity needed to the interaction, explaining, partly, how EB1 can act as an essential adaptor protein at the MT plus ends.

Due to the important role as a master regulator of protein-protein interactions at the MT plus ends the design of small molecule modulators to target EB1 is very attractive. This poses a challenge for two reasons, the first related with targeting a protein-protein interaction and the second related with the existence of IDRs in both proteins. Both subjects, intrinsically disordered regions in intermolecular interactions and protein-protein interactions, will be subsequently introduced.

## **1.2 Protein-protein interactions: a biological challenge for drug targeting**

### **1.2.1 Protein-protein interactions**

Protein-protein interactions (PPIs) play a central role in many biological processes.<sup>52</sup> With the recent developments in the area of genomics and proteomics, complete networks of PPIs within a cell (interactome) were identified, leading to major breakthroughs in understanding biological pathways, host-pathogen interactions and cancer development.

It is estimate that 650000 protein-protein interactions exist in human interactome.<sup>53,54</sup> PPIs play a key role in nearly every biological function<sup>55</sup>, regulating numerous cellular processes including signalling pathways, morphogenic pathways and complex molecular machines.<sup>56</sup> Furthermore, aberrant PPIs contribute to most disease states, representing a highly populated class of untouched potential targets for drug discovery.<sup>54</sup> Thus, it is of great importance to use and develop new technologies to target these PPIs by discovering PPI modulators, both as inhibitors or probes to understand these complex processes.<sup>54,57</sup>

Currently, three main classes of PPI modulators are known, antibodies, peptides and small chemical compounds, although nucleic acids and sugar molecules can also be used.<sup>54</sup> Antibodies have high specificity and are stable in human serum. However, they are not cell permeable and do not have oral bioavailability, making it difficult to use in therapeutics. Despite that, monoclonal antibodies *Avastin* and *Herceptin* are

currently in use.<sup>52</sup> Peptides usually suffer from poor metabolic stability, low bioavailability and can induce immune reactions.<sup>58</sup> However, miniature proteins still hold a considerable promise given their enhanced conformational and proteolytic stability.<sup>52</sup> The main advantage is due to its size (~20 residues), these peptides adopt a defined conformation improving its stability.<sup>57</sup> From a medicinal chemistry and drug development perspective, small molecules, which are generally cheaper to manufacture than antibodies and peptides and have good oral bioavailability, are still the preferred option.<sup>58,59</sup> The latter, small drug-like molecules, will be the main focus in this work.

PPIs can occur between two structured protein domains, a structured domain and a relatively short peptide, or between two peptide stretches. In many cases, additional weak contacts distant to the defined interaction area contribute to binding, thereby adding complexity and complicating the prediction of PPI characteristics.<sup>60</sup> Structural biology has shown that PPI interaction surfaces are generally large (~1500-3000 Å), flat<sup>61</sup>, and generally lack the grooves and pockets present at the surface of proteins that bind to small molecules.<sup>59</sup> Moreover, most contact surfaces in protein-protein interfaces involve amino acid residues that are not contiguous.<sup>59</sup> Unlike enzymes or G protein-coupled receptors (GPCRs), nature did not offer simple small molecules that can start a chemical discovery process and high-throughput screening (HTS) had not provided validated hits.<sup>62</sup> The chemical space of traditional small-molecule libraries deviates from that of PPI inhibitors, thus leading to low hit rates when applied in screening for PPI inhibitors. This stimulated the search for alternative strategies involving fragment-based screens or natural product inspired libraries that contain molecules with relatively high molecular weights and a large number of stereogenic centres.<sup>60</sup>

The investigation of numerous PPI interfaces revealed that a small subset of residues are the major contributors for the free energy of binding. These regions are generally called “hot-spots” and often overlap with structurally conserved regions and represent a common feature of PPI interfaces.<sup>57,60,61</sup> Hot spots tend to cluster at the centre of the interface, cover an area comparable to the size of a small molecule, be hydrophobic and show conformational adaptivity.<sup>62</sup> Using a hot-spot as the starting point for ligand identification is generally accepted as a useful strategy for the discovery of PPI small molecule modulators.<sup>52</sup>

Small molecules inhibit PPIs through at least three different mechanisms: orthosteric inhibition, allosteric regulation, and interfacial binding/stabilization. The first mechanism, orthosteric inhibition, involves direct competition against the interacting partners, binding to sites in the target proteins that overlap with the areas used for interacting with the partner proteins. The allosteric regulation relies on ligand binding induced changes in the target protein. These small-molecule ligands bind to sites distinct from the macromolecular interface inducing changes in either the static conformation or the dynamic properties of the target protein(s) and hindering the native interaction. Finally, an interfacial inhibitor binds to a pocket at the macromolecular interface, which sometimes is only transiently formed in a transition state, and locks the complex into a conformation that cannot be targeted by the natural ligand.<sup>63</sup>

Despite all the challenges and difficulties, in the past decade, an increasing number of successful cases of PPI inhibitors were reported.

#### 1.2.1.1 Success cases in modulating protein-protein interactions

The cytokine interleukine-2 (IL-2) has a key role in the activation of T cells and in the rejection of tissues grafts, having a considerable medical interest. A series of small molecules that bind to IL-2 were discovered at Sunesis Pharmaceuticals, where **SP4206** (Figure 1.12) showed to bind with high affinity. A fragment-based approach guided by X-ray structures and medicinal chemistry was used in this project since the structure of the IL-2-IL-2-R $\alpha$  complex was not known by that time.<sup>59</sup>

Another successful case in protein-protein interactions modulation is the B-cell lymphoma 2 (Bcl-2) family members. These are important regulators in apoptosis, and can form homo or heterodimers with other family members generating several combinations of pro-apoptotic and/or anti-apoptotic complexes. **ABT737** (Figure 1.12) binds to Bcl-X<sub>L</sub>, inhibiting its anti-apoptotic activity.<sup>59</sup> **ABT-263**<sup>64</sup>, a **ABT737** derivative has reached the clinical trials phase.<sup>53</sup> This group of compounds was discovered using a fragment-based NMR method known as structure–activity relationships (SAR) by NMR, and their properties were improved by using NMR-structure-guided medicinal chemistry.<sup>65</sup>

A third success case, the human protein double minute 2 (HDM2) has emerged as an excellent drug target for cancer treatment. It binds to the tumour-suppressor protein p53

and increases its degradation. A **benzodiazepinedione** and **Nutlin-3** (Figure 1.12) were found to disrupt the HDM2-p53 complex. Both these scaffolds were discovered through HTS and further optimisation of the found hits.<sup>59</sup>

These success stories elucidate some features about PPIs that can be useful in other drug discovery projects for PPI modulators. For instance, a common myth is that protein-protein interfaces are large and flat, often rigid and do not present cavities for small molecule binding. However, in all of the supramentioned cases the contact surfaces have some adaptability, and cavities that are not seen in structures of either the free protein or the protein-protein complex become available for binding.<sup>59</sup> A single crystal structure may not reveal potential binding modes for small molecules.<sup>54</sup> Eyrish & Helms<sup>66</sup> showed through molecular dynamics (MD) simulations that the protein surfaces of Bcl-X<sub>L</sub>, IL-2 and MDM2 form transient binding pockets that were not initially observed.

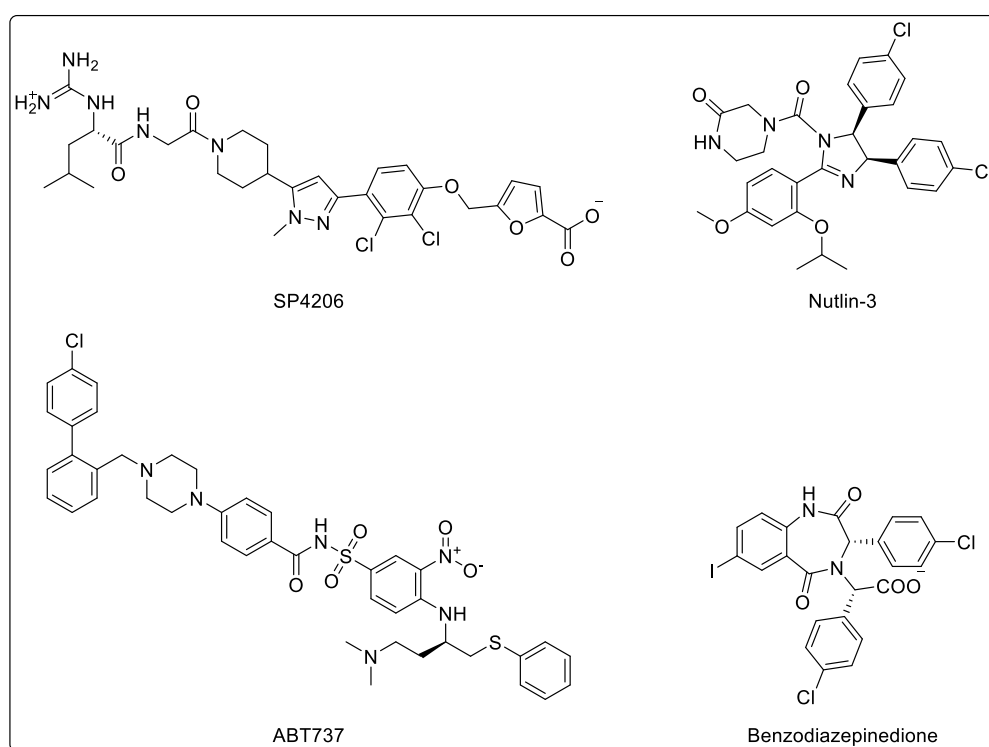


Figure 1.12 – Examples of small molecules that inhibit protein-protein interactions.<sup>59</sup>

All the examples described above included an empirical screening process, fragment screening or HTS.<sup>59</sup> Unfortunately, assays to investigate PPIs are usually difficult to apply to HTS. Therefore, structure-based drug design, *in silico* screening and other cheminformatics strategies need to be developed and applied to facilitate the development of PPIs modulators.<sup>58</sup> However, given that the type of interactions for

PPIs differ from those typical for enzyme and receptor drug targets, exploration of the chemical space defined by PPIs requires different approaches not biased by earlier drug discovery campaigns.<sup>61</sup> Commercial libraries of compounds used for screening are generally dominated by past drug discovery research into “druggable” targets, such as G-protein-coupled receptors or enzymes.<sup>59</sup>

A balance between potency and selectivity required for a PPI modulator and drug-like properties for oral administration need to be found in order to successfully target the vast range of PPIs. In other words, there is the need to find the chemical space of PPIs modulators, moving away from the libraries of compounds created primarily focused on other classes of targets.

#### ***1.2.1.2 Drug discovery in protein-protein interactions***

To tackle the challenge of finding small molecule inhibitors to target protein-protein interactions many approaches were used in the past, biophysical methods (X-ray crystallography and NMR), fragment-discovery approaches as SAR by NMR and tethering, HTS, and *in silico* approaches, such as structure- or ligand-based virtual screening, protein-protein docking, hot-spot and druggable pocket prediction tools.<sup>67</sup>

One of the most successful approaches applied to protein-protein interactions is fragment-based drug discovery. Fragment initial hits will have low affinity, but when tethered they allow novel areas of chemical space to be explored, creating diversity in chemical libraries.<sup>59</sup> However, few fragment-based campaigns of PPIs have been reported in the literature, let alone advanced into the clinic.<sup>68</sup> HTS of compound libraries has also been widely used.<sup>69</sup> Unfortunately, this strategy requires a large physical compound library, a suitable screening method and a validation method. This approach has been mostly used in pharmaceutical companies but some academic groups have been using it. Nevertheless, it requires not only a suitable system to be screened (not all biological systems may be analysed using the same techniques) but also a great investment in the latest technologies for compound screening, screening libraries, *etc.* This approach can be also time consuming. Alternative inexpensive and fast methods may be the rational drug design/identification through computational methods.

The first step in the rational drug design using computational methods, should be the creation and validation of a theoretical model, gathering information from the target

and known ligands.<sup>70</sup> Identifying *in silico* virtual hit molecules – virtual screening – is usually the next step.

### 1.2.1.2.1 Virtual Screening

The virtual screening of chemical libraries aims to identify new small organic molecules that are able to interact with a specific target, modulating its activity.<sup>70</sup> Virtual screening is the *in silico* analogue of biological screening, with its aim being to identify hit molecules from a chemical library using one or more computational procedures, to help decide which compounds to screen, synthesise or purchase.<sup>71</sup> There are two main strategies that can be used in virtual screening – ligand-based and structure-based design. The ligand-based design approach is based on knowledge of known molecules that bind that target (*e.g.* pharmacophore search). While, when the structure of the target is known molecular docking can be performed in a structure-based design approach<sup>72,73</sup> – Table 1.4.

Table 1.4 – General guidelines when making a decision on the method for computer aided drug design.

		Ligand(s)	
		✓	✗
Protein structure	✓	<b>Structure-based design</b> <ul style="list-style-type: none"> <li>• <b>Molecular docking</b></li> </ul>	<i>De novo design</i>
	✗	<b>Ligand-based design</b> <ul style="list-style-type: none"> <li>• <b>Similarity search</b></li> <li>• <b>Pharmacophore</b></li> <li>• <b>QSAR</b></li> </ul>	

Ligand-based design is based on the availability of pharmacologically relevant agents and their bioactivities, *i.e.* known ligands for a certain target. Approaches include similarity searching, a host of machine learning methods including quantitative structure activity relationships (QSAR) and pharmacophore mapping. If just a single molecule is known similarity searching is an option. If some ligands are known it is possible to identify a common pharmacophore, followed by a 3D database search. When a large number of ligands have been tested for a certain target and there are active and inactive structures then machine learning methods and QSAR approaches can be used.<sup>71</sup> Unfortunately, for EB1 there are no known active ligands, apart from the naturally occurring SxIP containing proteins. Therefore, the only ligand-based design method that can be applied to this system is the construction of a pharmacophore model based on the SxIP feature. A pharmacophore model is “an ensemble of steric and electronic features that is necessary to ensure the optimal



supramolecular interactions with a specific biological target and to trigger (or block) its biological response".<sup>74</sup> From the practical point of view, pharmacophores can be used to screen millions of high quality compound structures within a reasonable amount of time, particularly when approximations such as rigid pharmacophore fitting procedures are used.<sup>75</sup> The positioning of key features like hydrogen-bonding and hydrophobic groups is summarised in a pharmacophore representation. It can be used as a template to select the most promising candidates from a library or as a filter before applying a structure-based virtual screening method, so that only 1–10% of the initial database has finally to be docked.<sup>76</sup>

Structure-based virtual screening methods offer means to directly identify novel compounds that complement the target protein surface; these methods are not limited by the requirement for template compound(s) implicit to ligand-centric (mimicry) approaches.<sup>77</sup> Docking is the most used tool for structure-based virtual screening.<sup>76</sup> The docking process involves the prediction of ligand conformation and orientation (*pose*) within a specific binding site.<sup>73</sup> Docking algorithms *pose* small molecules in the binding site, being complemented by scoring functions that evaluate the interaction between the molecules and the target – *scoring*.<sup>73</sup>

The success of structure-based drug design is well documented; it has contributed to the introduction of ~50 compounds into clinical trials and to numerous drug approvals.<sup>78</sup> In the last years, diverse research groups proposed an integration of ligand- and structure-based strategies to increase the success of virtual screening processes. On one side pharmacophore information of ligands have been incorporated in docking studies, and on the other side docking studies have been carried out finding suitable conformations, inside the target binding cavity.<sup>79</sup> These successes were achieved by using a combination of methods, molecular docking, pharmacophore search, application of filters, amongst others.<sup>80</sup>

To illustrate the success of computational methods in drug discovery, *Aggrastat* (Figure 1.13), an antiplatelet drug that belongs to glycoprotein IIb/IIIa inhibitors class is one of the first commercialized drug whose discovery was influenced by virtual screening methods (in this case pharmacophore search).<sup>81</sup>

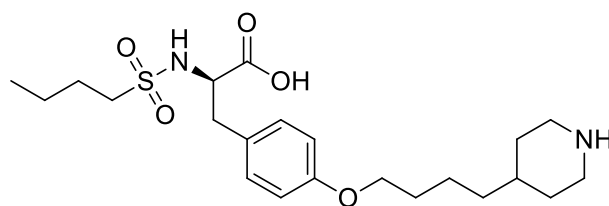


Figure 1.13 - Aggrastat, an antiplatelet drug of the glycoprotein IIb/IIIa inhibitors class whose discovery was influenced by virtual screening methods.

### 1.2.1.3 The chemical space of PPI modulators

The physicochemical properties of the novel chemical entities identified in virtual screening have become the subject of intense scrutiny from lead discovery to drug candidate. The selection of compounds with acceptable drug-like properties is often performed as a post-dock processing.<sup>56</sup> In this section the focus will be on molecular properties important for drug-like compounds and how the traditional drug like properties compare with the PPI modulators, as many reports have found that small molecules targeting PPIs often fall outside the chemical space of the current drugs.<sup>62,82</sup> The meaning of “drug-like” is dependent on mode of administration. The original Rule of 5 (Ro5)<sup>83</sup> deals with orally active compounds and defines four simple physicochemical parameter ranges: molecular weight (MW)  $\leq 500$ ;  $\log P \leq 5$ ; H-bond donors  $\leq 5$  and H-bond acceptors  $\leq 10$ ; these rules are associated with 90% of orally active drugs that have achieved phase II clinical status.<sup>84</sup> If a compound fails two or more of the Ro5 there is a high probability that the compound will display poor oral availability. However, passing the Ro5 is no guarantee that a compound is drug-like.<sup>84,85</sup>

In many cases, at least modifications of Lipinski’s rule-of-five have been allowed to account for the larger size and higher hydrophobic content of protein–protein interfaces. In fact, by analysing some physicochemical properties of known small molecule modulators of PPIs it is evident that some of the identified compounds violate one or two Lipinski’s rules, in particular, MW and/or  $\log P$  and exhibit solubility issues.<sup>67</sup>

Traditionally, affinity is the first aspect considered for hit selection and optimization. However, affinity alone can be misleading as it is often found to be linked with molecular size. Thus a focus on affinity leads to a bias towards a selection of bigger compounds. In addition, optimization of affinity during subsequent stages of drug discovery typically leads to a further increase in molecular weight.<sup>86</sup> The idea of ligand

efficiency (binding energy/non-hydrogen atoms) has recently emerged as a useful guide to optimize fragment and lead selection in the discovery process. Moreover, the ligand efficiency coefficient used to analyse protein–ligand interactions during the course of drug discovery programs has been revisited for PPI modulators and estimated to be 0.24 kcal/mol *per* heavy atom.<sup>58,62</sup> This is lower than most kinase inhibitors (0.3–0.4 kcal/mol) but on the same order of magnitude of many protease inhibitors (0.25–0.35 kcal/mol). Consequently, a PPI modulator with a  $K_d$  of 10 nM is expected to have a molecular weight of approximately 645 g.mol<sup>-1</sup> when regular orally available drugs, most of the time, are below 500 g.mol<sup>-1</sup>.<sup>58</sup> There are other important issues for drug likeness, including lack of reactive functionality (except in prodrugs), cell permeability, and, for central nervous system compounds, brain/blood partitioning. An acceptable level of solubility is also critical to permit dissolution and absorption; virtually all drugs have aqueous solubility above 10<sup>-6</sup> M ( $\log S > -6$ ).<sup>78</sup>

There have been several attempts to find the chemical space of PPI modulators. A study presented by Morelli and co-workers in 2011 suggest an adaptation of Lipinski's Ro5 to a "rule of four".<sup>53</sup> After statistical analysis of the 39 inhibitors these authors report average values for molecular weight ( $547 \pm 154$  g.mol<sup>-1</sup>, thus MW > 400 g.mol<sup>-1</sup>), ALogP ( $3.99 \pm 2.37$ , thus ALogP > 4), number of rings ( $4.44 \pm 1.02$ , thus #Rings > 4) and number of hydrogen bond acceptors (HBA) ( $6.62 \pm 2.60$ , thus #HBA > 4). Figure 1.14 shows a comparison between Ro5 and rule of four.

Analysis of PPI focused libraries has suggested that, in addition to the features formerly presented, PPI modulators often present branched structures and have high content in multiple bonds, *i.e.* have a more tri-dimensional shape around a central scaffold.<sup>58</sup> The preference towards a tri-dimensional and branched shape may be related with the fact that the hot spots on the target protein are commonly composed by two to three sub-pockets<sup>62</sup> and a branched structure has more chances to fill in these sub-pockets.

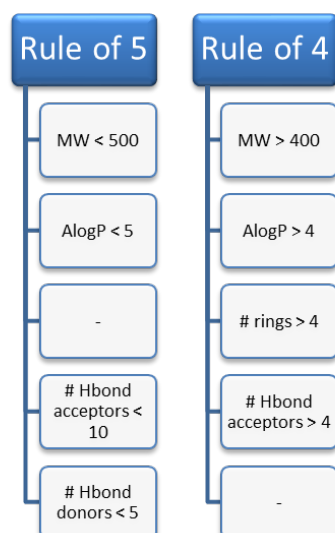


Figure 1.14 – Comparison of Lipinski rule (rule of 5)<sup>83</sup> and the values reported for the definition of a generic profile for PPI inhibitors (rule of 4) reported by Morelli et al.<sup>53</sup>

When looking to protein-protein interfaces one of the interesting features of hot spots is their functional and structural adaptivity. Many proteins function by binding to multiple partners. An important point is that these proteins tend to use the same hot spot, which adapts to present the same residues in different structural contexts.<sup>87</sup>

### 1.2.2 Disorder and dynamics in protein-protein interactions

For many protein-protein interactions, the apparent complementarity between the two surfaces involves a significant degree of protein flexibility and adaptivity.<sup>87</sup> As presented previously the EBH domain of EB1 does not only bind to IDRs containing SxIP motifs, but itself contains an IDR that becomes structured upon complex formation (section 1.1.5).

The binding process in dynamic/disordered regions is of great interest from both mechanistic and functional standpoints. Traditionally, there are two models focused on flexibility to describe binding processes, induced fit and conformer selection. The popular notion of induced fit assumes a passive mutual adaptation, but it does not explain specific recognition if the two protein structures are not complementary to start with. The opposing idea of conformer selection suggests recognition between the two bound conformations that are postulated to occur within the diverse structure ensembles of the two free proteins.<sup>88</sup> Either the conformational change precedes the

binding (conformational selection) or occurs following the binding (induced fit). An alternative model that mixes both concepts have been proposed by Grünberg and co-workers in 2004, where some conformations among the structure ensembles of unbound proteins are more prone to recognition than others.<sup>89</sup>

Therefore, there might be binding-site conformations that are well-suited to small-molecule binding yet are not visible in a single crystal structure.<sup>87</sup> Solution methods for analysis of these complexes with dynamic regions seem to be fundamental to understand the underlying mechanisms of binding.

### 1.2.3 Screening techniques in drug discovery applied to protein-protein interactions

As pointed out before, many of the recent clinically approved drugs have originated from HTS campaigns, through the use of microtiter plate-based assays against libraries of small molecules.<sup>90</sup> When it comes to hit validation, target-based biochemical assays have proven their value in PPI drug discovery projects, but cell based assays have significantly become more important. Despite that, target validation and understanding the underlying mechanisms of action can only be understood through target based assays.<sup>90</sup> A summary of the assay methods, protein/reagent requirement and associated assay cost estimations are shown in Table 1.5. Post-screening hit validation studies are employed to gain an understanding of the mode of action of the molecules. The low-throughput assays described are considered as being the most appropriate for this as they provide information such as  $K_d$ ,  $k_{on}$ ,  $k_{off}$ , identification of which protein within the PPI complex they bind to as well as their stoichiometry. Cell-based assays are also highly attractive during the hit validation phase as a positive outcome (*e.g.*, inhibition of the PPI) assumes that the investigated compounds can penetrate the cell membrane and are not cytotoxic.<sup>90</sup>

Some of the gold standard assay formats to monitor PPIs have a low-throughput and advances are being made to them so as to increase their capacities.<sup>90</sup> A recent study based on an integrated biophysical approach for fragment screening has shown that the use of several different biophysical techniques was essential for the validation of hits proceeding from the thermal shift analysis. Whereas thermal shift provided a fast way to screen the entire library, NMR allowed the subsequent validation of the resulting hits. Thermal shift provided a >50% rate of false positives that were later verified not

to bind by NMR; however it significantly enriched the hit rate of the secondary NMR screen. Most fragments validated by NMR and isothermal titration calorimetry (ITC) were shown by X-ray crystallography to bind, making these techniques ideal for compound screening and hit validation.<sup>91</sup>

Table 1.5 - Screening techniques applied in drug discovery assays. This table was originally published by Gul & Hadian in 2014<sup>90</sup>

Technique	Throughput	Assay type	Protein/reagent requirement	Assay cost
Isothermal titration calorimetry	Low	Biochemical	High	Low
Surface plasmon resonance	Low	Biochemical	Low	Low/medium
Microscale thermophoresis	Low	Biochemical	Low	Low/medium
ELISA	Low	Biochemical	Low	Low
Pull-down	Low	Biochemical	Medium/high	Low/medium
NMR and X-ray crystallography	Low	Biochemical	High	Medium/high
Amplified luminescent proximity homogeneous assay screen	High	Biochemical	Low	High
Time-resolved Förster resonance energy transfer	High	Biochemical	Low	High
Fluorescence polarisation	High	Biochemical	Low	Low
Förster resonance energy transfer/bioluminescence	High	Cell-based	Low	Low
Biomolecular fluorescence complementation	High	Cell-based	Low	Low
<i>In situ</i> proximity ligation assay	Medium/high	Cell-based	Low	Medium/high
Co-immunoprecipitation	Low	Cell-based	Medium/high	Low/medium

In this project the HTS approach was replaced by a more directed virtual screening approach and therefore there is no need of a screening technique that can be applied to HTS. The focus of this section will be on the two main techniques used for compound screening – solution NMR and ITC.

### 1.2.3.1 Nuclear magnetic resonance (NMR)

The principles of this technique will not be described herein as this is a technique with a very wide range of applications and a description of the principles that can be applied to compound screening and protein analysis seem to be more appropriate in this context.

NMR can provide a lot of information on a protein-protein or protein-ligand interaction. It can be used as a screening tool (with  $K_d$  determination possible even for weak interactions), structure determination of a complex and dynamic, kinetic and thermodynamic aspects of the interaction.<sup>92,93</sup>

In recent years, NMR has become a valuable screening tool for the binding of ligands

to protein targets, and has the key advantages of being able to detect and quantify interactions with high sensitivity without requiring prior knowledge of protein function.<sup>94</sup> An advantage of this technique is its versatility, as one can either observe the resonances of the ligand or the resonances of the target.<sup>92</sup> The observation of the ligand resonances is usually applicable to ligands in the low-affinity range (high nanomolar to high millimolar range). This is commonly used when either the protein is too large to be observed by NMR (> 100 kDa), not available with the desired <sup>15</sup>N or <sup>13</sup>C labelling pattern from *in vitro* production, or it aggregates in solution at high concentrations.<sup>92</sup> Screening based on the observation of the target protein resonances relies on detection of chemical shifts changes of the protein upon titration of the ligand.<sup>92</sup> In the standard experiment, one needs <sup>15</sup>N-labelled protein plus an unlabelled ligand, which can be a small molecule or another macromolecule. The ligand is titrated into the protein, monitored at each stage of the titration by acquiring a <sup>1</sup>H, <sup>15</sup>N-HSQC spectrum. The <sup>1</sup>H, <sup>15</sup>N-HSQC spectrum is the easiest to assign, it is sensitive, and signals are usually well resolved, particularly in comparison with the <sup>1</sup>H, <sup>13</sup>C-HSQC. The experiment detects just one signal *per* amino acid corresponding to the backbone HN-group (excepting proline), and a small number of side-chain signals (asparagine, glutamine and tryptophan, and sometimes, depending on the solution conditions, arginine and histidine). The chemical shifts are very sensitive to structural changes and can be measured very accurately, meaning that almost any genuine binding interaction will produce chemical shift changes in the target.<sup>94</sup> If the structure and backbone assignment of the protein is known these chemical shift changes will indicate which residues are being affected upon binding. Moreover, the shape of the titration curve (chemical shift *versus* concentration of ligand) can often be fitted straightforwardly to obtain a value for the dissociation constant of the ligand,  $K_d$ . NMR titrations can, therefore, directly provide both a  $K_d$  value and a binding site from the same set of measurements.<sup>94</sup> The calculation of  $K_d$  values by NMR will be detailed in Chapter 5 (section 5.1.2).

In addition to the mapping of the binding site, one of the first steps in any study of protein-ligand interactions by NMR is to establish to which region of exchange the spectrum corresponds (or, more correctly, the resonances of interest, since different resonances can, of course, show different exchange behaviour).<sup>93</sup>

For a protein  $P$  binding reversibly to a ligand  $L$  at a single site,



characterised by rate constants for forward and back reactions of  $k_{on}$  and  $k_{off}$ , the dissociation constant  $K_d$  is equal to  $[P][L]/[PL]$ , where  $[P]$ ,  $[L]$  and  $[PL]$  represent the concentrations of free protein, free ligand and complex (Equation 1.1)

$$K_d = \frac{[P][L]}{[PL]} \leftrightarrow K_d = \frac{k_{off}}{k_{on}}, \text{ Equation 1.2}$$

$K_d$  can be thought of as the concentration of the free ligand when half of the binding sites are occupied. The forward and back rates are given by  $[P][L]k_{on}$  and  $[PL]k_{off}$  respectively – Equation 1.2. At equilibrium the forward and back rates are equal, implying that the dissociation constant  $K_d$  is also equal to  $k_{off}/k_{on}$ . When exchange is fast, *i.e.* when  $k_{off}$  is much greater than the chemical shift difference, then the signals in the  $^1H, ^{15}N$ -HSQC spectrum will move smoothly from their position in the free spectrum to those in the bound spectrum, with the frequency of the signal at any titration point being the weighted average of free and bound shifts – Figure 1.15, first panel. This is usually characteristic of a weak interaction. When the exchange rate is slow on the chemical shift timescale, or in other words when  $k_{off}$  is significantly slower than the difference in Hz between the chemical shifts of free and bound protein, then as the ligand is titrated in, the free signal gradually disappears and the bound signal appears, the intensities of the two peaks reflecting the concentrations of free and bound protein – Figure 1.15, third panel. Slow exchange is usually a good indicator of strong binding. If the exchange rate is similar to the shift difference, then signals broaden and shift at the same time – intermediate exchange – Figure 1.15, middle panel.

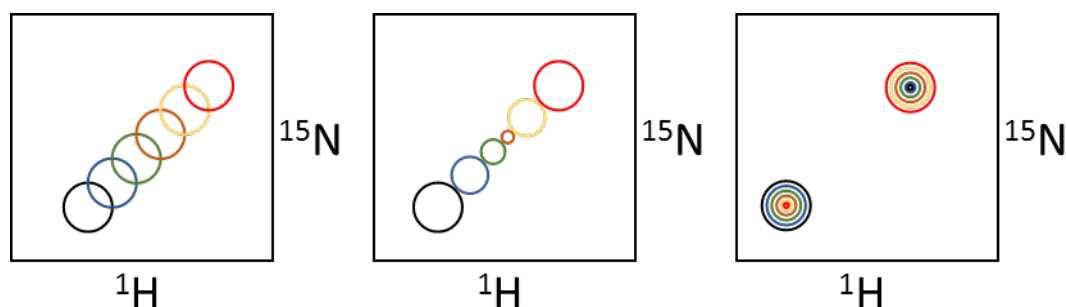


Figure 1.15 – Schematic representation of the effect of exchange regimes on signals in  $^1H, ^{15}N$ -HSQC spectra in a ligand titration into a target  $^{15}N$  labelled protein. First panel is an example of a fast exchange regime, middle panel intermediate exchange and third panel shows a system in slow exchange.

This usually leads to broadening or disappearance of signals, especially in the



intermediate points of titration.<sup>94,95</sup>

Analysis of the spectral changes described is sufficient for ligand screening. However, NMR also offers a possibility of structural analysis of a complex in solution. This usually requires stable isotope enrichment of the sample,  $^{15}\text{N}$  and  $^{13}\text{C}$ , and sometimes  $^2\text{H}$  labelling for large proteins ( $> 30$  kDa).<sup>95</sup>

Data acquisition and resonance assignment is the first step of structure determination. For that, a wide range of 2D and 3D experiments need to be acquired. Three-dimensional triple resonance experiments that correlate hydrogen ( $^1\text{H}$ ), nitrogen ( $^{15}\text{N}$ ) and carbon ( $^{13}\text{C}$ ) are fundamental for backbone and sequential assignment. Side-chain assignment is usually achieved by using 3D HCCH-total correlation spectroscopy (TOCSY) and HCCH-correlational spectroscopy (COSY). Once all backbone and side-chain assignments are complete, distance restraints are derived from nuclear Overhauser effect spectroscopy (NOESY) data. Intermolecular NOESY contacts (between the two elements that form the complex) can be acquired using isotope-*filtered* NOESY experiments. These experiments remove coherences of protons attached to  $^{15}\text{N}$  or  $^{13}\text{C}$ , leaving only signals from protons of the unlabelled ligand detected in one or more dimensions. In contrast, *edited* experiments select signals of proton attached to  $^{15}\text{N}$  or  $^{13}\text{C}$ . The isotope-filtered NOE experiments are recorded in addition to the conventional 3D  $^{13}\text{C}$ - or  $^{15}\text{N}$ -edited NOESY experiments, and comparison with the latter allows separation of inter- from intramolecular contacts.<sup>95</sup> All the NOE data is converted into distance restraints to perform the structure calculation. Here, two problems arise, the NOE assignment and the calculation of structures that fulfil the conformational restraints derived from the NOE assignment. Molecular dynamics, minimisation and simulated annealing have been successfully applied to structure calculation. However, there are no robust methods for NOE assignment and this step is still very time consuming and laborious due to the large number of assignment possibilities, peak overlap and potential artefacts in the spectra.<sup>95,96</sup> Programs for NOE automated assignment were developed, for instance CANDID/CYANA, AUTOSTRUCTURE or ARIA (Ambiguous Restraints for Iterative Assignment)<sup>96</sup>. The focus will be on ARIA as it was the program used in this project for automated NOE assignment. ARIA introduced a very important concept in NOE assignment – the use of ambiguous distance constraints. The majority of NOEs cannot be assigned unambiguously from chemical shift information alone and the

treatment of some distance restraints as ambiguous facilitate the assignment of NOEs as the same peak can be assigned to different spin systems. In a later stage (iteration), ARIA should be able, at least in theory, to refine the assignments accordingly with the lower energy structures obtained from previous iterations.<sup>97</sup> ARIA assigns NOE cross-peaks by first deriving all possible assignments for each peak by matching a list of chemical shifts. Peak volumes are converted into distance restraints using the isolated spin pair approximation, which relates the volume to the inverse sixth power of the distance between the two interacting spins (Equation 1.3).

$$V_{ij} = \alpha d_{ij}^{-6}, \text{ Equation 1.3}$$

where  $V_{ij}$  is the volume and  $d_{ij}$  is the distance.<sup>95,96</sup> Distance restraints are then calculated based on the cross-peak assignment and Equation 1.3. These restraints are subsequently used to calculate structures using Crystallography & NMR System (CNS)<sup>98</sup> that employs a simulated annealing refinement method. Other restraints, not derived from NOE data can be introduced into the calculations to facilitate the process and give higher quality models. Examples are hydrogen bonds,  $J$ -couplings, residual dipolar couplings, disulphide bridges and dihedral angle restraints. Obtaining a correct initial fold at the outset of a *de novo* structure determination can be challenging because the powerful structure-based filters used for the elimination of erroneous cross peak assignments are not yet operational at that stage. It is of great help for the initial phase of the algorithm if the user can supply a limited number of already assigned long-range distance constraints, meaning the user still has to provide some assignments in order to get the correct protein fold.<sup>97</sup> An overview of the whole process for structure determination using solution NMR can be found in Figure 1.16.

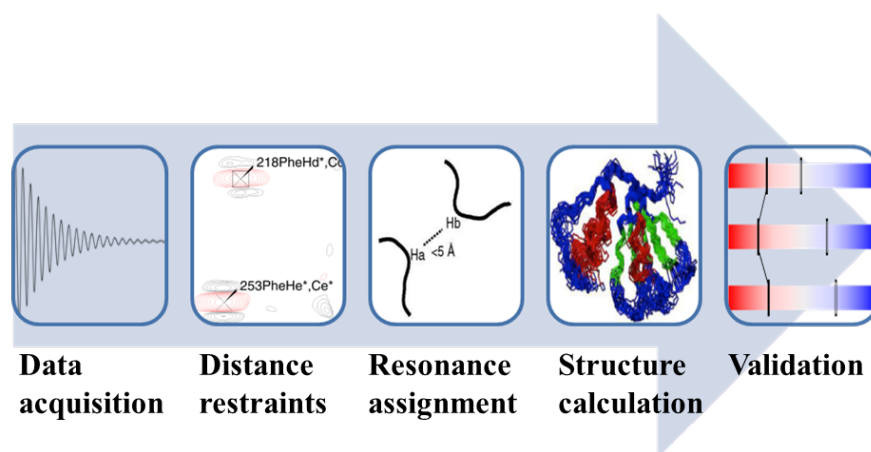


Figure 1.16 – General overview of the process for structure determination using solution NMR.

There is still a long way until structure determination via solution NMR reaches the automation existing for more established methods such as X-ray crystallography. However, the importance of the information acquired using NMR for characterisation of complexes in solution is undeniable.

### ***1.2.3.2 Isothermal titration calorimetry (ITC)***

The ITC method measures energy (heat) changes that are associated with the binding reaction of two components. It is a method that can directly determine the binding affinity ( $K_a$ ) and enthalpy ( $\Delta H$ ), and using these values it is possible the calculation of free Gibbs energy ( $\Delta G$ ) and entropy ( $T\Delta S$ ). This is achieved by measuring stepwise changes in the heat changes during the course of a titration experiment. The underlying stoichiometry of an interaction can also be determined using this method.<sup>90,99</sup> Therefore, ITC has found widespread applicability in the study interaction analysis.<sup>99</sup>

The experiment takes place inside a reaction cell containing one of the binding partners. Subsequently, the other binding partner (protein or ligand) is injected step-wise into the reaction cell (titration) until the required excess of the titrant is reached. The heat changes are determined by comparing the reaction cell to the reference cell.<sup>90</sup>

Enthalpic and entropic factors can either contribute favourably or unfavourably to  $\Delta G$  (Equation 1.4) resulting in the four possible modes: (i)  $\Delta H > 0$ ,  $\Delta S < 0$ ; (ii)  $\Delta H < 0$ ,  $\Delta S < 0$ ; (iii)  $\Delta H < 0$ ,  $\Delta S > 0$ ; and (iv)  $\Delta H > 0$ ,  $\Delta S > 0$ . Only modes (ii–iv) yield negative  $\Delta G$  values, leading to binding. While many protein-ligand binding events are driven by enthalpic factors, in some cases entropy can contribute favourably towards a negative change in free energy and, thus, result in binding.<sup>100</sup>

$$\Delta G = \Delta H - T\Delta S, \text{ Equation 1.4}$$

The change in  $\Delta H$  is determined by a variety of interatomic forces, including electrostatic, van der Waals, and hydrogen-bonding interactions, and the entropic contribution  $\Delta S$  represents the change in the size of the conformational space available to the overall system, including the protein, ligand, and solvent molecules.

Most of the entropically favourable interactions are hydrophobic in nature because of large positive solvation entropy changes on reduction of the exposed hydrophobic surfaces. Before interacting, the protein and ligand are each solvated separately. Upon binding, which in this context usually involves the burial of hydrophobic surfaces,

many water molecules from the separate hydration shells surrounding the protein and ligand are freed into bulk solvent. This process increases the number of conformational states of the water molecules and, thus, of the system as a whole, overcoming the loss of entropy due to restraining the molecule and binding site.<sup>100</sup>

This method is immobilization-free (solution method and closer to native conformations) and has the advantage of not requiring any labelling of any investigated protein. ITC can be used to detect interactions between two proteins or between a protein and a small molecular weight ligand. However, sample consumption can be relatively high when using this technique and therefore it has limited applicability for proteins that are difficult to produce in relatively large amounts.<sup>90</sup> Another limitation of the ITC experiment is having upper/lower limits of binding affinity of ~10 nM and ~1 mM respectively. These limits are related with the ideal concentrations for the macromolecule. High affinity interactions (low  $K_d$ ) should be studied at low concentrations: however, the minimum concentration that will typically cause a confidently measurable heat change for a 1:1 interaction is about 10  $\mu$ M. Low affinity interactions (high  $K_d$ ) should be studied at high concentrations, but the concentration that can be used may be limited by availability or solubility of the sample molecule.<sup>99,101</sup> The development of competition (or ligand displacement) ITC experiments has led to binding parameter determination for very strong (or weak) binders.

### 1.3 Research Aims

EB1 is a key protein at the MT plus ends protein network, nevertheless, no small molecules ligands have been identified or designed to target EB1. Targeting EB1 using small molecules could bring immense benefits for a better understanding of its role *in vivo* or even stabilise/destabilise its binding to other molecules – protein-protein interaction modulators, ultimately making it “druggable”.

The main aim of the project is to identify a chemical scaffold based on the SxIP motif that can target the EBH domain of EB1. This scaffold can be used for the design of molecules that can modulate the interaction of EB1 with SxIP proteins. To achieve this goal, intermediate aims were defined. These, include the definition of one or more virtual screening approaches to identify hit molecules and binding characterisation of

the identified hits using biophysical screening techniques, such as NMR. The understanding of how SxIP ligands interact with EB1, identifying which regions promote the specificity and which regions contribute to the affinity of the interaction will be also covered as this will help to characterise the native interaction and this information can be used for subsequent drug design campaigns - Figure 1.17.

These aims will be covered in five chapters, where chapters 2, 3, 4 and 5 will be focused mainly in the two first aims (virtual screening and ligand binding characterisation) with Chapter 5 covering the characterisation of the native interaction. Chapter 2 will cover the initial attempt to identify EB1 ligands by using a molecular modelling approach including pharmacophore search based on the known SxIP motif and docking using a crystal structure of EB1. Compounds identified using this method will be screened for EB1 binding.

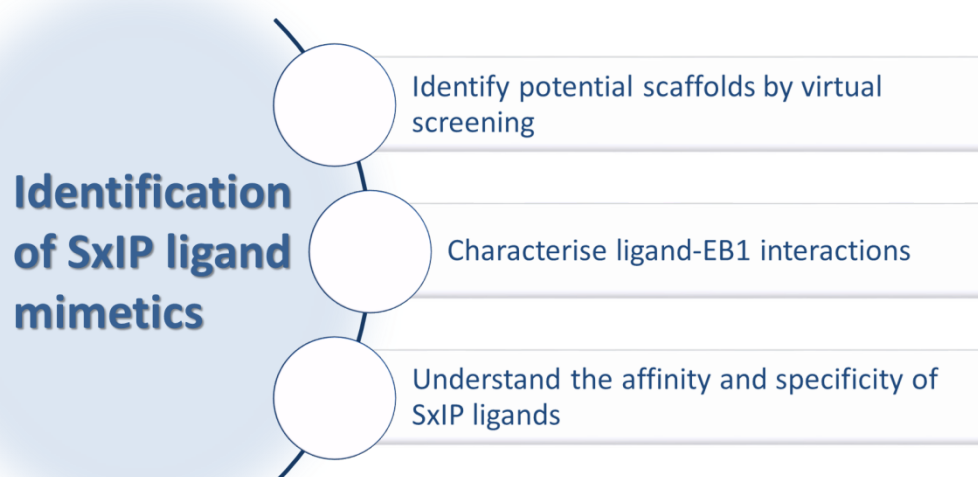


Figure 1.17 – Main aim defined for this project and sub-aims needed to achieve the main goal.

The next chapter (Chapter 3) includes structural characterisation, using solution NMR, of EB1 in the free form and bound to one of the ligands identified in the previous chapter. The aim of this chapter is to understand how these compounds interact with EB1. Differences between these structures will be described and this information used for another attempt to find small molecule inhibitors to target EB1 – Chapter 4, extending the contact region beyond the SxIP motif.

Chapter 5 is a detailed study of short length peptides based on a SxIP protein native sequence and their interactions with EB1. The aim of this chapter is to understand the

## *Introduction*

contributions of different regions in the interaction between SxIP proteins and EB1. The final section of this chapter is the identification of higher affinity peptide sequences that can, in the future, be used to build a new pharmacophore and identify EB1 small molecule ligands.

## Chapter 2. IDENTIFICATION OF NOVEL SMALL MOLECULE INHIBITORS

Finding novel small molecule inhibitors that can bind to EB1 is the main aim of this chapter. As described in the introduction – Chapter 1, finding small molecule modulators for protein-protein interactions is considered one of the current challenges in drug discovery. Therefore, a highly comprehensive and targeted approach will be used.

The crystal structure of a complex formed between the C-terminal of EB1 lacking the last eight C-terminal residues (EB1c $\Delta$ 8) and a 30 residue peptide derived from the C-terminal of human MACF2 (MACFp1)<sup>31</sup> is available on the PDB<sup>102</sup> with code 3GJO. With the complex structure available it is possible to use both ligand and structure-based design approaches for identification of small molecule modulators. The natural ligand, MACFp1, allows to search for compounds that can establish similar interactions with the target, EB1c $\Delta$ 8, whereas the target itself permits further refinement of a potential binder through molecular docking. Both pharmacophore-based and docking based methods have proven to be successful in virtual screening projects; however none of the approaches is clearly superior.<sup>103</sup> Therefore, the use of both approaches can be beneficial. Solution NMR will be the biophysical technique to screen the selected compounds.

### 2.1 Virtual screening methods applied to identify hit molecules to target EB1

The outcome of the screening process is highly dependent on the type of virtual library created.<sup>104</sup> As described before the chemical space of PPI inhibitors has not been completely defined. There have been efforts to generate PPI focused libraries, for example, the 2P2I<sup>53,105,106</sup> and TIMBAL<sup>107</sup> databases. In order to enhance the chemical diversity a PPI focused library was not used, instead ZINC database<sup>108,109</sup> was utilised to build the compound libraries for the structure-based design. With more than 35 million purchasable molecules, ZINC, is the largest database of commercially available compounds for virtual screening.<sup>108</sup> One of the reasons ZINC database was chosen is related to the fact that requirements as suitable for organic synthesis and/or

purchasability are very important.<sup>56,110</sup> The second reason is related to a tool embedded with this database – ZINPharmer<sup>55</sup>. ZINCPharmer provides tools for constructing and refining pharmacophore hypotheses directly from molecular structure. Pharmer identifies hydrophobic, hydrogen bond donor/acceptor, positive/negative ions and aromatic pharmacophore features, searching in ZINC database for conformers that can satisfy these features. The concept of hot-spots in PPIs was previously presented. Despite there being no non-peptide compounds known to bind to EB1, the four residue peptide SxIP has been frequently reported as fundamental for the interaction with the EBH domain of EB1. Moreover, as shown before this motif targets a highly conserved hydrophobic pocket. By using this knowledge on the native interaction it is possible to reduce the number of compounds to be docked and analysed through structure-based design methods.

In terms of molecular docking a clear propensity towards GOLD<sup>111-118</sup> (Genetic Optimisation for Ligand Docking) has apparently emerged with respect to other docking software for PPI drug discovery projects.<sup>67</sup> This may be due to two reasons. GOLD uses a genetic algorithm (GA) that can be applied to more flexible systems. It is known that during the physical binding, both the ligand and the protein adapt their conformations to each other (flexibility). As a consequence, docking algorithms should handle the flexibility of both ligand and protein. Almost all docking programs perform flexible ligand docking while the receptor is kept rigid. One of the exceptions is GOLD, which apply some flexibility to the protein during the docking, through active site chain rotations and, in a more global level, minimizations.<sup>76</sup> This is done by using a GA. The GA works by using an evolutionary strategy in exploring the full conformational flexibility of the ligand with partial flexibility of the protein. It also satisfies the fundamental requirement that the ligand must displace loosely bound water on binding.<sup>112</sup> The other reason may be the fact GOLD presents four scoring functions, allowing the score to a certain binding pose to be attributed based on a wide range of different parameters. The four scoring functions will be subsequently presented.

Goldscore has a Van der Waals treatment of clash and dispersion terms and uses a crystal structure derived treatment of hydrogen bonding and metal terms

Chemscore is an empirical scoring function parameterised from 82 complexes of known binding affinity. It has a lipophilic-lipophilic contact area term, a geometrically



constrained hydrogen bond term and a term to penalise excessive flexibility.

ASP (Astex Statistical Potential) is a knowledge-based scoring derived from the Protein Data Bank (PDB)<sup>102</sup>

ChemPLP (Piecewise Linear Potential) is the most recently introduced scoring function. This treats neutral and repulsive contacts with a piecewise linear potential (PLP). This simple potential has both an attractive and repulsive part for neutral contacts and solely a repulsive part for neutral contacts and solely a repulsive part for anti-complementary contacts (donor-donor, metal-donor and acceptor-acceptor). The Chemscore hydrogen bonding term is used for hydrogen bonds and the Chemscore internal energy term is also used. ChemPLP is fast to calculate in comparison to Goldscore (benchmarked at 23 seconds for ten times repeat dockings at default settings, compared to Goldscore, benchmarked at 90 seconds).<sup>119</sup>

Among the available GOLD scoring functions, Chemscore is known to be suited for binding sites for which a significant part of the ligand affinity can be ascribed to hydrophobic interactions, as is often the case in protein–protein interfaces, while the complementary Goldscore is preferred for polar binding sites.<sup>67</sup> Different scoring functions evaluate the binding using different parameters, having different deficiencies and strengths. Sometimes, it is desirable to combine different scoring functions in order to obtain better predictions. Thus, a recent trend, consensus scoring, has been introduced in molecular docking studies. Consensus scoring gathers information from different scoring functions improving the probability of identify “true” ligands.<sup>73</sup> Consensus scoring is recognized to increase hit rates by reducing the number of false positives and the errors in the scores/ranks and it should be considered as the most reasonable strategy when groping in the dark.<sup>67</sup> In addition to a good virtual screening protocol, it is often important to prevent investing time and resources in compounds that will most likely fail at a later stage of the drug discovery process. One of the greatest challenges in the virtual screening remains the simultaneous prediction and optimization of both binding affinity and pharmacokinetic properties. *In silico* methods can also be used to predict and analyse molecular properties of the hit molecules, in a fast and inexpensive way. The prediction of properties related to drug-likeness, as taking onboard ADME (absorption, distribution, metabolism and excretion) consideration early in pre-clinical development, may help to avoid costly late-stage pre-clinical and clinical failures.<sup>78</sup>

The reasons for this failure have been widely described and may include poor pharmacokinetic properties, lack of efficacy or selectivity, toxicity and unexpected adverse effects.<sup>120</sup> This cannot be totally avoided, although there is knowledge based on past experience that can help to flag these undesirable properties. These properties can be conflicting with other desirable features such as structure based design predictions or chemical diversity. To overcome these conflicting properties one can use a multi-objective analysis, where a pool of equivalent optima are obtained for all the objectives (properties).<sup>120,121</sup>

## **2.2 Identification of small molecules based on SxIP motif**

It has been widely reported that a diverse group of +TIPs, the SxIP proteins, contain conserved SxIP motifs in intrinsically disordered regions, enriched in basic, serine, and proline, residues. The SxIP motif binds, specifically, to the C-terminal domain of EB1 – EBH domain.<sup>31,49</sup> The crystal structure confirms the most prominent contacts involve Ser5477, Ile5479 and Pro5480, positions 1, 3 and 4 of the SxIP motif. Ser5477 forms an extensive network of hydrogen bonds with highly conserved Arg222, Glu225, Gln229 and Tyr247 through a water molecule – Figure 2.1A.<sup>31</sup> Lys5478 is within a salt bridge distance (~4 Å) to Asp257; however the electron density is poorly defined for both side chains, indicating a dynamic region.<sup>31</sup> More importantly, Ile5479 and Pro5480 are buried within a hydrophobic cavity formed by Lys220', Leu221 and Leu246 and delimited by the aromatic rim at the bottom, Phe216', Tyr217' and Phe218, followed by Arg222, Glu225, Tyr247, and Ala248 – Figure 2.1B.<sup>31,43</sup>

SxIP motif targets the hydrophobic pocket and surrounding areas of EB1c and it does seem a good starting point for the design of small molecules that can bind to EB1 and therefore block its interaction with other proteins.

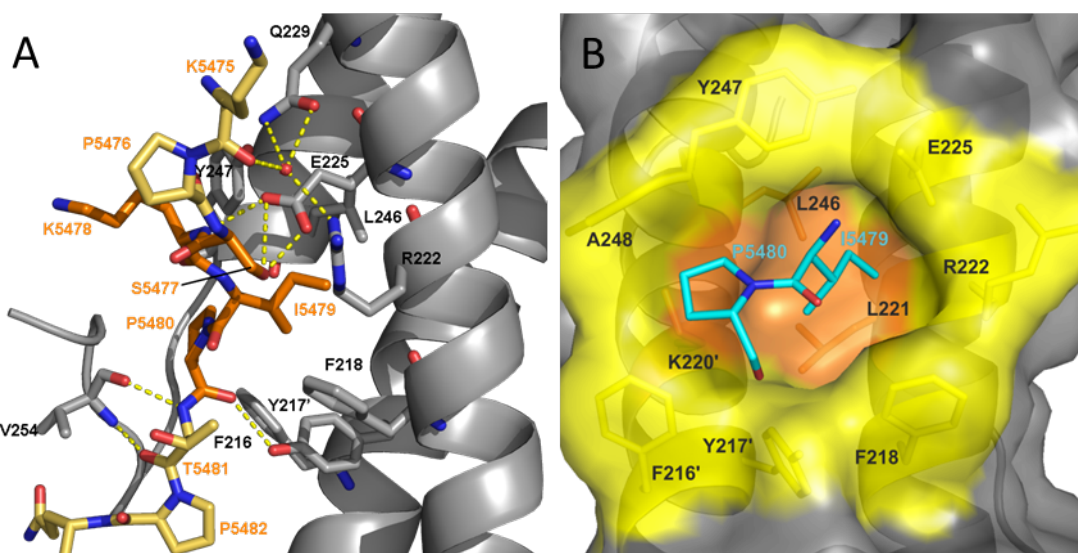


Figure 2.1- A - Close-up view of the crystal structure of the complex EB1c $\Delta$ 8-MACFp1. SxIP motif is coloured orange EB1 represented as cartoon ribbon (grey). MACFp1 is represented as sticks, with the SxIP motif coloured in orange and the remaining residues coloured in yellow. (PDB code 3GJO)<sup>15</sup> The yellow dashed lines represent hydrogen bonds. B – IP motif (cyan) of MACFp1 and hydrophobic pocket of EB1c, formed by Lys220', Leu221 and Leu246 (coloured in orange) and delimited by Phe216', Tyr217' and Phe218, followed by Arg222, Glu225, Tyr247, Ala248 (coloured in yellow). C-terminus residues (Thr249-Gly260) were removed for a better visualisation.

### 2.3 Ligand-based design - Pharmacophore search

A pharmacophore is the ensemble of steric and electrostatic features of different compounds which are necessary to ensure optimal supramolecular interactions with a specific biological target.<sup>122</sup> In other words, a pharmacophore is an abstract concept that describes the common steric and electrostatic complementarities of bio-active compounds with the target.<sup>123</sup> As referred before, ZINCPharmer was used as a tool for pharmacophore definition and database filtering and the SxIP motif was used as the starting point to define a pharmacophore model. Besides the type of the interaction (hydrogen bond acceptor/donor, hydrophobic, positive/negative ion and aromatic), other key information can be incorporated into the pharmacophore model such as the three-dimensional location of the interaction, if an interaction is directional (such as hydrogen bonding) and also its spatial orientation.<sup>123</sup> ZINCPharmer considers that hydrogen bond acceptors/donors must be within 4 Å of a hydrogen bond donor/acceptor and a hydrophobic feature must be within 6 Å of at least three hydrophobic features of the receptor. The direction of the hydrogen bonding is

represented, by arrows, while the size of the hydrophobic regions is controlled by the radius of the respective sphere - Figure 2.2.<sup>55</sup> In this case, since no receptor structure was uploaded for the search, radius values and bond directionality were independent of the receptor.

Due to the polar interactions observed for Ser5477 (described in the previous section), two hydrogen bond acceptors were defined – Table 2.1. Three hydrophobic centroids were defined for Ile5479 and Pro5480 since these residues clearly make hydrophobic contacts. Lys5478 was not utilised for the definition of a pharmacophore model since it seems to make contacts with a more dynamic region and the aim at this stage is to find small molecules that can interact mainly with the hydrophobic cleft formed by the two monomers of EB1. In addition, backbone amide for Ile5479 and carbonyls for both Ile5479 and Pro5480 were included as hydrogen bond acceptor and donors, respectively – Figure 2.2 and Table 2.1. The inclusion of both hydrophobic contacts plus hydrogen bond contacts allow for a well-fitted molecule binding to the pocket, stabilised by hydrogen bonds.

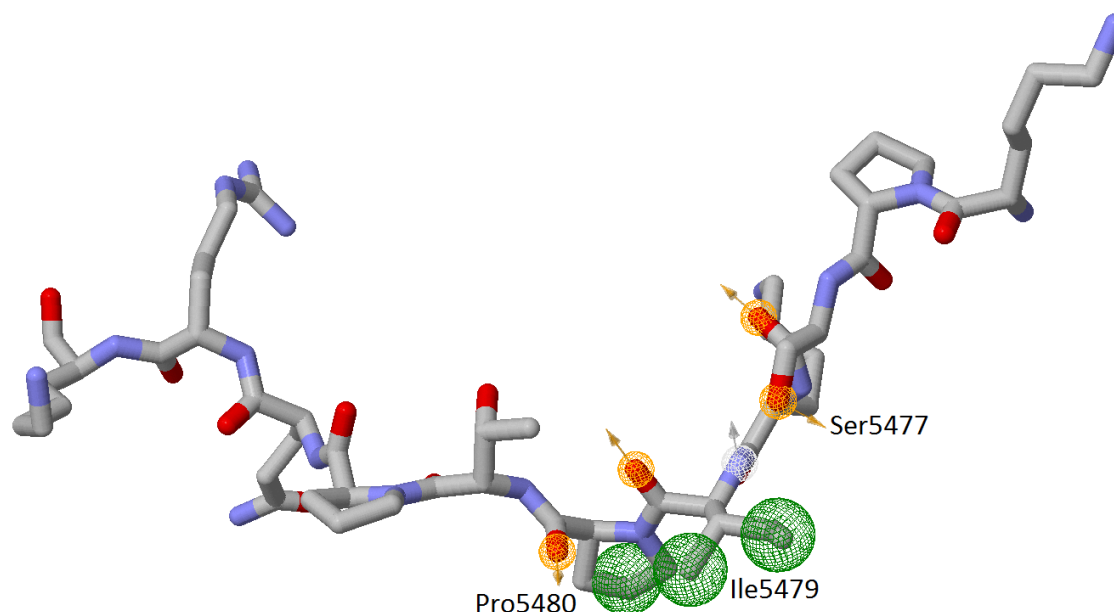


Figure 2.2 - Representation of the pharmacophore points found for Ser5477, Ile5479 and Pro5480 of the MACFp1 peptide. All pharmacophore features are shown in spheres, being the hydrogen bonding acceptors showed as orange mesh, hydrogen bonding acceptors in white mesh and hydrophobic as green mesh. Orange and grey arrows indicate the direction of the hydrogen bond donor/acceptor, respectively.

Table 2.1 - Pharmacophore points found based on Ser5477, Ile5479 and Pro5480

Pharmacophore Class	Coordinates			radius	residue
	x	y	z		
H donor	-16.20	-30.47	21.17	0.50	Ile5479
H acceptor	-13.21	-33.20	21.86	0.50	Ser5477
H acceptor	-13.64	-30.55	34.37	0.50	Ser5477
H acceptor	-15.25	-28.40	30.36	0.50	Ile5479
H acceptor	-16.20	-25.55	29.38	0.50	Pro5480
Hydrophobic	-16.92	-26.78	32.96	1.00	Ile5479
Hydrophobic	-19.01	-27.81	29.06	1.00	Pro5480
Hydrophobic	-16.39	-28.30	35.09	1.00	Ile5479

Since a ligand by itself does not provide information about the nature of an interaction, the result is not a true pharmacophore. Instead, the pharmacophore derived from a single ligand structure should be seen as a 3D similarity search.<sup>55</sup>

The search for molecules within the ZINC database containing all these pharmacophore points, with a molecular weight smaller than 500 g.mol<sup>-1</sup>, gave no hits. Thus a systematic search was done, removing one, two and three pharmacophore points, yielding a total of 71 possible combinations. The result was that 40006 molecules were found to fit at least five of the defined pharmacophore points. Since the number was high, another filter, Root Mean Square Deviation (RMSD) smaller than 0.5, was applied retrieving 3933 molecules, which is a much more reasonable number to perform molecular docking. Filtering by RMSD restricts the hits to those that have the best overall geometric match to the query.<sup>55</sup> Since ZINCPharmer offers different conformations of the same molecules, repeated molecules were removed, retaining 3060 molecules for the docking studies.

## 2.4 Molecular Docking studies

### 2.4.1 Molecular Docking method validation and general considerations

Before performing the molecular docking studies method validation is needed to check that docking can predict accurately the binding modes of an interaction.

*Pose selection* is a commonly used method for validation of docking methods and scoring functions. A compound with a known conformation and orientation, typically

from a crystal structure, is re-docked and the results given (poses) are evaluated in terms of similarity with the known conformation.<sup>124,125</sup> Usually a RMSD lower than 2 Å between the experimental (crystal structure) and the calculated poses is considered a successful validation.<sup>76,114</sup> To validate the docking method the peptide fragment of MACFp1 – SKIP was used since the main aim of the molecular docking studies in this project was to make a virtual screening of small molecules, with a molecular weight and size closer to the SKIP motif rather than the 11 residue peptide presented in the crystal structure – KPSKIPTPQRK, *i.e.* ~443 g.mol<sup>-1</sup> instead of ~1279 g.mol<sup>-1</sup>.

Additionally, the importance of probing the hydration state of a protein binding pocket in the un-bound state has been demonstrated and should be a standard element of structure-based design.<sup>126</sup> For this reason, the pose validation was also performed in the presence of a water molecule in the binding site. The crystal structure shows the presence of a water molecule that mediates hydrogen bond formation between MACF2 and EB1.<sup>31</sup> When water molecules are known or assumed to play a role in protein-ligand recognition, the most common strategy is to perform separate docking runs in parallel, *i.e.*, one in the absence of water molecules and a second in the presence of one or more water molecules.<sup>116</sup>

Hydrogen bonds are the most important specific interactions in biological processes. Their geometries follow strict rules and therefore this is part of the information that can be extracted from crystal structures.<sup>126</sup> Due to the importance of these interactions, and based on the information retrieved by the crystal structure 3GJO, the following hydrogen bond constraints were introduced in GOLD – Arg222, Glu225, Leu246, Tyr247, Val254 from chain A and Tyr217 from chain B.<sup>31</sup>

Table 2.2 - The average and standard deviation calculated based on 10 docking solutions for the SxIP motif

Rotatable bonds	Fixed		Not fixed	
	Score	RMSD (Å)	Score	RMSD (Å)
GoldScore	71.4 (±0.1)	0.25 (±0.02)	70 (±8)	2.3 (±0.9)
ChemPLP	68.00 (±0.06)	0.48 (±0.02)	77 (±5)	3 (±2)
ChemScore	34.6 (±0.2)	0.5 (±0.2)	22 (±2)	3 (±2)
ASP	35.5 (±0.2)	0.45 (±0.08)	36 (±1)	5 (±2)

Table 2.3 - The average and standard deviation calculated based on 10 docking solutions for the SxIP

motif, with a H<sub>2</sub>O molecule and hydrogen bonding constraints

Rotatable bonds	Fixed		Not fixed	
	Score	RMSD (Å)	Score	RMSD (Å)
Goldscore	37.0 (±0.6)	0.40 (±0.03)	47 (±6)	3 (±2)
ChemPLP	31.1 (±0.1)	0.47 (±0.03)	57 (±7)	3 (±2)
Chemscore	-5.5 (±0.1)	0.30 (±0.05)	-9 (±4)	6 (±2)
ASP	-4.02 (±0.08)	0.33 (±0.04)	-3 (±8)	7 (±2)

Goldscore and ChemPLP were found to have a better performance when posing the natural ligand – SKIP - for this model. The first molecular docking study, without any constraints, gave a smaller RMSD for Goldscore, 0.25 Å when compared with ChemPLP, 0.48 Å. The same situation is observed when constraints are applied to the docking method, with Goldscore showing a slightly smaller RMSD value, 0.40 against 0.47 Å obtained for ChemPLP. In this case the difference is much smaller and considering the standard deviation values, not significant. Based on both results it was decided to choose Goldscore as the scoring function for this study.

The definition of the binding site was based on a radius of 6 Å around the ligand included in the crystal structure Table 2.4. EBH domain of EB1 is a homodimer and therefore the binding site is formed by residues belonging to both monomers – chain A and chain B.

Table 2.4 - EB1 active residues used for GOLD for the docking of small molecules inhibitors. These residues were defined using a 6 Å radius around the ligand present in the crystal structure.

Chain A					Chain B				
Arg214,	Phe218,	Leu221,	Arg222,	Glu225,	Asp209, Glu213, Phe216, Tyr217, Lys220				
Leu226,	Gln229,	Leu246,	Tyr247,	Ala248,					
Thr249,	Asp250,	Glu251,	Gly252,	Phe253,					
Val254,	Ile255,	Pro256,	Asp257						

With this binding site definition the area of interaction was not restricted to the region where the SxIP motif binds but expands considerably further. This is an important feature since it is known that in order to enhance specificity and affinity, additional interactions, beyond those present in a single individual “hot spot” are required. Nearby residues that do not meet the criteria of a “hot spot” may also play an important, if not essential role in the interaction.<sup>127</sup>

## 2.4.2 Docking of virtual library obtained from ZINCPharmer

Effective consideration of key ligand-target interactions, such as hydrogen bonds, and other environmental factors during docking, such as target flexibility, metal ions, and water molecules, can enhance the docking-based virtual screening performance.<sup>128</sup>

Thus, four molecular docking studies were performed, evaluating the 3060 molecules and including different variants, like hydrogen bonding constraints and a presence of a water molecule. There could be several potential advantages to including water molecules in a protein-ligand docking program. First, if the compound interacts with the water molecule, including it could improve the predicted binding mode. Several studies have been reported in the literature where parallel dockings were done in the absence of water molecules and in the presence of some key water molecules. Some authors have reported significant improvements in docking performance when water molecules were included, whereas others found that including water molecules had little effect on the quality of the dockings.<sup>116</sup> Regarding the hydrogen bonding constraints, GOLD allows the definition of key hydrogen bonds. Molecules that form these defined bonds will score higher. In this study the residues chosen from EB1 to form key hydrogen bonds with the ligand were: from chain A, Arg222, Glu225, Gln229, Val254 and Tyr247 and Tyr217, from chain B, in agreement with what was shown by Honnappa and co-workers.<sup>31</sup>

In the first docking study no hydrogen bonding constraints or water molecule were included. The second and third docking studies included hydrogen bonding constraints and a presence of a water molecule, respectively, while the fourth docking study included both features - Table 2.5.

Table 2.5 - Summary of the considerations, presence of water molecules or definition of key hydrogen bonding, made for each docking study.

	H <sub>2</sub> O molecule	H bond constraints
1	No	No
2	No	Yes
3	Yes	No
4	Yes	Yes

At this stage, due to the large number of molecules the poses and scores obtained were not directly inspected. Instead, an analysis of the whole dataset was performed, using an ensemble of techniques described subsequently.



## 2.5 Screening and selection of hit molecules

### 2.5.1 Consensus scoring

Due to the inherent limitations of the scoring functions, and taking advantage of GOLD providing four different scoring functions, the best docked conformer of each compound was re-evaluated using other scoring functions – this procedure is known as *Consensus Scoring*.<sup>129</sup> Compared to single scoring procedure, *Consensus Scoring* reduces the number of false positives, improving the hit-rates.<sup>129</sup> This was one of the conclusions from the results of an extensive computational study, of docking collections of three-dimensional structures into three different enzymes of pharmaceutical interest: p38 MAP kinase, inosine monophosphate dehydrogenase, and HIV protease.<sup>130</sup>

For this purpose, docking poses obtained using Goldscore were rescored using the alternative scoring functions: ASP, Chemscore and ChemPLP.

### 2.5.2 Ligand Efficiency

Large molecules can form many hypothetical interactions in binding sites and therefore have the tendency to generate better scores than smaller compounds.<sup>131</sup> Ligand efficiency is a measurement of the binding energy *per* atom of a ligand to its binding partner, such as a receptor or enzyme. In the case there is no information on binding affinities or binding energies, but there is a docking predicted score that can be used instead. By normalizing the energy score obtained from the dockings studies a docking *Ligand Efficiency* (dLE) can be calculated:

$$LE = \frac{\text{score}}{n \text{ of heavy atoms}}, \text{ Equation 2.1}$$

### 2.5.3 Docking descriptors calculation

The protein-protein interfaces have large contact areas and are typically dominated by steric and hydrophobic interactions, although there are also electrostatic interactions and hydrogen bonds.<sup>56</sup> The latter types of interactions may be considered desirable to include as they facilitate the selection of more specific inhibitors which bind through these specific interactions.<sup>56</sup> Using GoldMine, a tool provided with GOLD that allows

the analysis of large quantities of docking information, some docking descriptors, including information of the predicted hydrogen bonds between the molecules and the protein, were calculated.

Between the calculated descriptors, there was the number of total hydrogen bonds and the number of specific hydrogen bonds. This second descriptor was defined based on the hydrogen bonds of the following residues Arg222, Glu225, Gln229, Tyr247, Val254 from chain A and Tyr217 from chain B, in agreement with what was shown by Honnappa and co-workers<sup>31</sup> and with the constraints defined for two of the four docking studies. Another descriptor, *hydrogen bond efficiency*, was calculated based on the ratio between the total number of hydrogen bonds and the specific hydrogen bonds.

Using Knime<sup>132</sup>, and the *rule engine* node, when the total number of hydrogen bonds was the same than the number of specified hydrogen bonds, the hydrogen bond efficiency value given was 1. When these features did not match, the *rule engine* node would retrieve 0.

#### **2.5.4 Multi-parameter optimization**

Many accomplishments in modern drug discovery are hindered by the lack of consideration of multiple molecular properties in the early stages of lead identification and optimization. Indeed, one of the common causes for lead compounds to fail in the later stages of drug discovery is the lack of consideration of multiple objectives (*e.g.*, ADMET properties) at the early stage of optimization of candidate compounds. Improving the pharmacological profile of a candidate molecule requires the optimization of numerous, often competing objectives (*i.e.*, biological or chemical properties), to discover the few improved molecules that represent the best compromise of the multiple criteria is determinant for a successful drug.<sup>133,121</sup> Pareto-based methods are capable of optimizing numerous properties simultaneously.<sup>121</sup>

Using the *Pareto Ranking* tool offered by Knime, a multi objective analysis was performed in order to select the compounds which had a balance between scoring values from the docking and molecular properties. The 3060 docked molecules were, therefore, ranked using the following properties:

## Consensus Scoring and Ligand Efficiency

### Drug-like properties<sup>83,134</sup>

A<sub>LogP</sub><sup>135</sup> ≤ 5

LogD ≤ 4

180 ≤ MW ≤ 500

Molecular solubility – **the higher value the better**

Molecular Polar Surface Area ≤ 140

Num of H bond acceptors ≤ 10

Num of H bond donors ≤ 5

Num of rotatable bonds ≤ 12

### Hydrogen bond

Total H bond – **the higher value the better**

Specific H bond – **the higher value the better**

H bond efficiency – **a match between *total H bond* and *specific H bond* was ranked higher**

After ranking the 3060 docked molecules using the above described *Pareto Ranking* approach, the top ranked 100 molecules were chosen to move forward in the virtual screening process.

### 2.5.5 Top ranked 100 molecules – selection of candidate compounds to synthesis and biological assays

To confirm the good binding predictions given by the first docking studies, it was decided to perform a molecular docking study for the selected 100 compounds. Since the number of molecules is smaller, the best ten solutions instead of one were saved in the solution file. If multiple solutions are generated for a ligand it is possible to calculate the RMSD of atom positions between each pair of solutions, using GoldMine. A smaller RMSD value is indicative that the docking prediction is very similar for the ten solutions and consequently gives more strength to the predicted poses and scores.

In addition, a visual inspection of the poses given by the 10 different solutions was performed. The hydrogen bonds formed between the molecules and the residues were inspected. These 100 molecules were shown to form hydrogen bonds with the following residues: Arg214, Arg222, Glu213, Gly252, Leu221, but more frequently with Ala248, Gln229, Glu225, Leu246, Lys220, Phe218, Pro256, Tyr217, Tyr247 and

Val254. This is a good indicator since some of these residues are frequently described in the literature<sup>27,31,43</sup> as components of the EB1 C-terminal binding site. However, a critical analysis also shows that these residues are within the defined binding site for the molecular docking studies.

The aqueous solubility of a compound is an important property that influences both the bioavailability as well as the magnitude of many ADME properties. However, solubilisation is a complicated process influenced by lipophilicity, hydrogen bonding formation, crystal packing and counterion.<sup>110</sup> With this in mind, for the selection of the hit compounds for synthesis, biophysical and biological testing, solubility (logS) was a key factor. Within the 100 molecules list, only molecules with a solubility value higher than - 4 were selected leading to the selection of only nine molecules.

Observing the way in which the molecules fit in the binding pocket and possible interactions with the target it was decided to select three molecules **ZINC12677264**, **ZINC08389070**, and **ZINC71025726**.

Table 2.6. ZINC12677264 showed to be the best ranked compound overall, especially based on the score/ligand efficiency ranking, rank #1, and hydrogen bonds formed, rank #3. The Goldscore fitness score is also the highest of these three candidate molecules. ZINC08389070, the third candidate molecule, has a very similar scaffold to ZINC12677264, performing better in terms of drug-like properties (rank #2), including slightly better solubility value and RMSD (rank #4) and indicating the calculated pose is very consistent. ZINC71025726, has the lowest Goldscore fitness value of the three compounds, but overall a very balanced position in all the ranks. ZINC12677264 and ZINC08389070 are result of the same pharmacophore definition, whereas ZINC71025726 results from a different ensemble of pharmacophore features. Both searches included three hydrophobic features for Ile5479 and Pro5480 and a hydrogen acceptor feature for Ile5479. However, the search that retrieved ZINC71025726 included a hydrogen acceptor feature based on the carbonyl of the Pro5480 rather than a hydrogen donor feature based on Ile5479 – Table 2.7 and Table 2.8.

Identification of small molecule inhibitors

Table 2.6 - Summary of rankings and molecular properties for the three candidate molecules resulting from the first virtual screening approach to target EB1. Final rank corresponds to the Pareto analysis using rank 1, score and ligand efficiency, rank 2, drug-like properties, rank 3, hydrogen bonds and rank 4, RMSD between the obtained solutions.

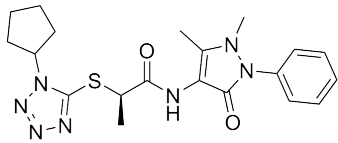
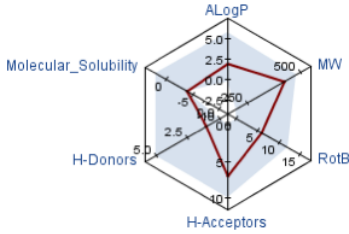
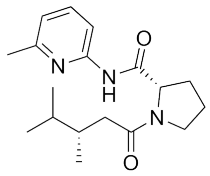
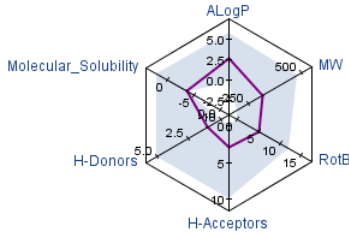
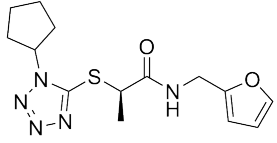
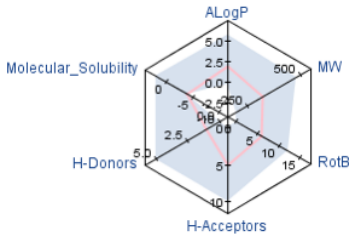
Molecule name	Final Rank	Rank 1	Rank 2	Rank 3	Rank 4	Solubility	Goldscore Fitness	Structure	Molecular properties
ZINC12677264	1	5	7	2	10	-3.992	62.3789		
ZINC71025726	4	8	4	3	6	-3.785	55.3401		
ZINC08389070	9	11	6	5	4	-3.96	55.9363		

Table 2.7 - Pharmacophore points used for the search that retrieved ZINC71025726.

Pharmacophore Class	residue
H acceptor	Ile5479
H acceptor	Pro5480
Hydrophobic	Ile5479
Hydrophobic	Pro5480
Hydrophobic	Ile5479

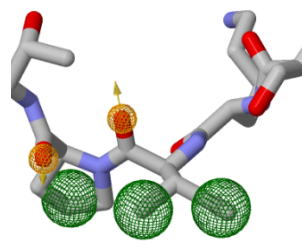
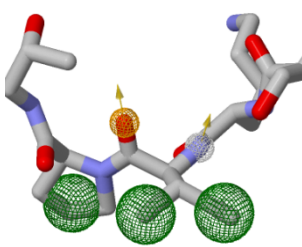


Table 2.8 - Pharmacophore points used for the search that retrieved ZINC08389070 and ZINC12677264.

Pharmacophore Class	residue
H donor	Ile5479
H acceptor	Ile5479
Hydrophobic	Ile5479
Hydrophobic	Pro5480
Hydrophobic	Ile5479



Despite these differences all molecules are based on pharmacophore features of Ile5479 and Pro5480. The conformation of ZINC71025726 is the most similar to the natural ligand when compared with other two molecules – Figure 2.3, left hand side panel. The RMSD value is, as expected, smaller, 0.275 Å when compared with ZINC08389070, 0.475 Å and ZINC12677264, 0.476 Å.

Regarding the docking poses (right hand side panel of Figure 2.3) one can observe that for ZINC71025726 the best scored pose is not much different than the conformation obtained from ZINCPharmer. On the other hand, for the other hits the binding poses are significantly different from the result obtained from ZINCPharmer. Notable features from the highest scored docking poses are, that the aliphatic chain of ZINC71025726 seems to emulate the  $\beta$ -branched side chain of isoleucine and the proline ring present in this molecule superimposes with Pro5480.

For ZINC08389070 the isoleucine hydrophobic chain is replaced by the cyclopentyl ring, and the Pro5480 replaced by a methyl group. The replacement of the isoleucine by a ring can still yield the expected hydrophobic contacts, but constricts the rotation, since the methyl groups were replaced by a more rigid group. The methyl replacing the proline does not seem to be a disadvantage also since the area where the proline of MACFp1 sits is quite small – Figure 2.1B.

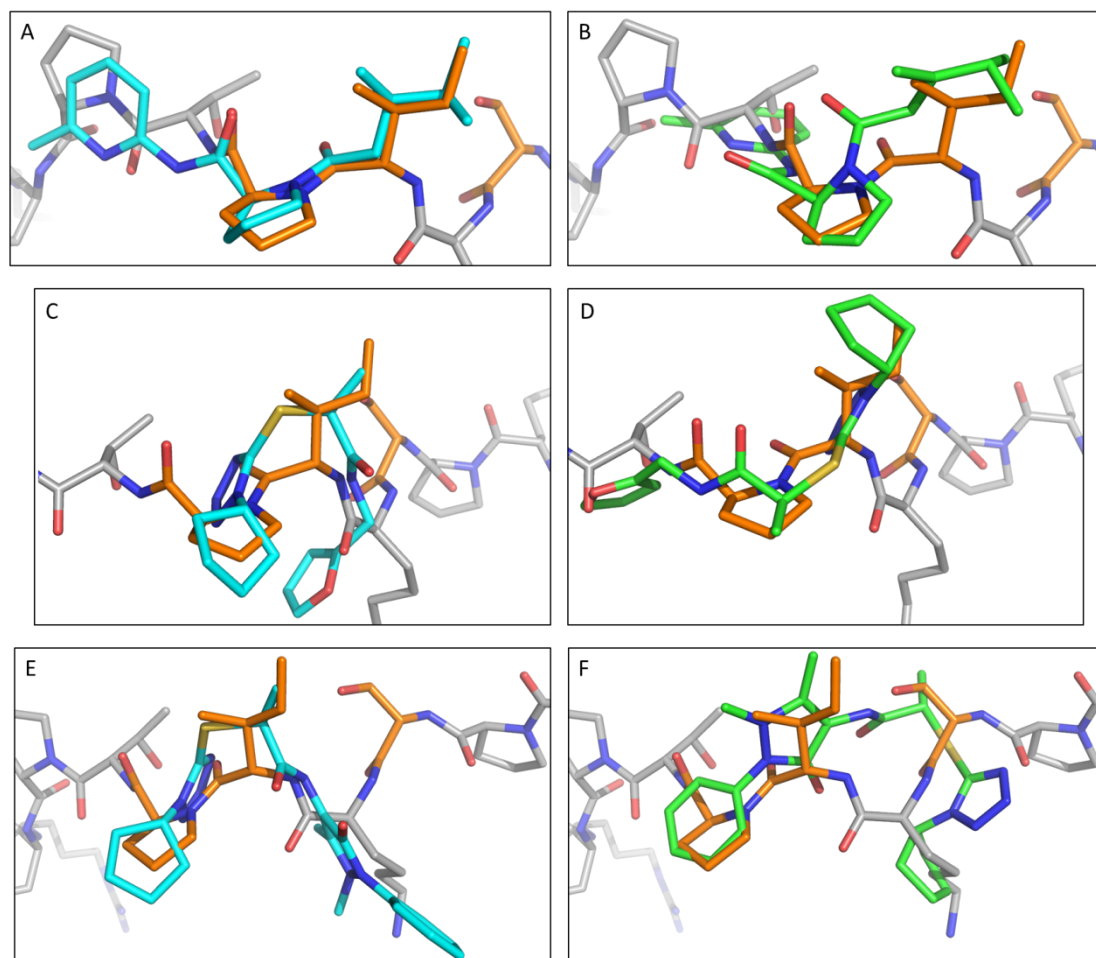


Figure 2.3 - ZINCPharmer results versus Gold docking poses for the three hit molecules. On the left hand side, the results of the search based on the defined pharmacophore model in ZINCPharmer for A – ZINC71025726, C – ZINC08389070 and D – ZINC12677264. On the right hand side best scored pose for B - ZINC71025726, D – ZINC08389070 and F – ZINC12677264. The crystal structure of 11MACFp1 is showed in grey, with Ser5477, Ile5479 and Pro 5480 highlighted in orange.

Finally, ZINC12677264 best scored docking pose does not provide the same interesting features as the previous compound, despite the scaffold being very similar. However, if analysing the ten best docking poses for this compound four out of ten, present a similar conformation to ZINC08389070, and an aromatic  $\pi$ - $\pi$  stacking interaction with Phe218 – Figure 2.4.

Since, ZINC12677264 was not commercially available at the time, it was decided to test ZINC08389070, and ZINC71025726 and find alternatives to ZINC12677264. Despite being the best ranked molecule, ZINC12677264, has a very similar scaffold to ZINC08389070, and using this approach it is possible to increase the chemical diversity of the tested molecules.



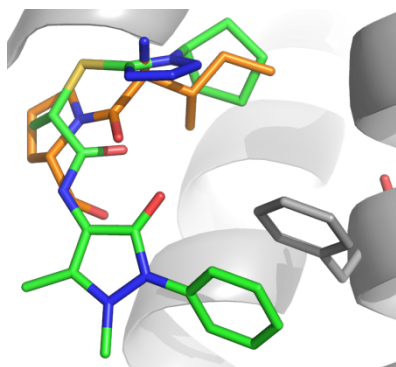


Figure 2.4 - Alternative pose for ZINC12677264. The cyclopentane ring and the methyl group emulate Ile5479 and Pro5480, respectively. The aromatic ring seems to establish aromatic  $\pi$ - $\pi$  stacking interaction with Phe218.

### 2.5.6 Alternative compounds to ZINC12677264

As stated before, ZINC12677264 was not commercially available and its scaffold is very similar to ZINC08389070. Chemical diversity is desirable at an initial stage of the drug discovery projects, therefore it was thought to select the best looking and consistent pose for ZINC12677264, build a new pharmacophore model and search for alternatives in ZINC database. In addition of being the best ranked molecule, ZINC12677264, shows a consistent pose where the aromatic ring seems to be in the right position to form  $\pi$ - $\pi$  stacking interactions with Phe218 – Figure 2.4.

#### 2.5.6.1 ZINCPharmer search

Using the pose shown in Figure 2.4, one can observe hydrogen bond formation between one of the nitrogens in the tetrazole ring and Gln229 and between the NH and Glu225, in addition to the aromatic  $\pi$ - $\pi$  stacking interactions with Phe218. Two different pharmacophore models were built based on these interactions – Figure 2.5. For the first one, A, only six hits were obtained; therefore, it was decided to remove two of the hydrogen bond acceptor groups within the tetrazole ring, leaving only the one that clearly seems to establish contact with Gln229 – B. For the latter 166 hits were retrieved. It is worth mentioning that the receptor crystal structure with the code 3GJO, was also uploaded into ZINCPharmer so the interaction pharmacophore will be automatically generated based also on the receptor features, *e.g.* hydrogen bond directionality.<sup>55</sup> Duplicated molecules were removed, giving a final list with 150 molecules.

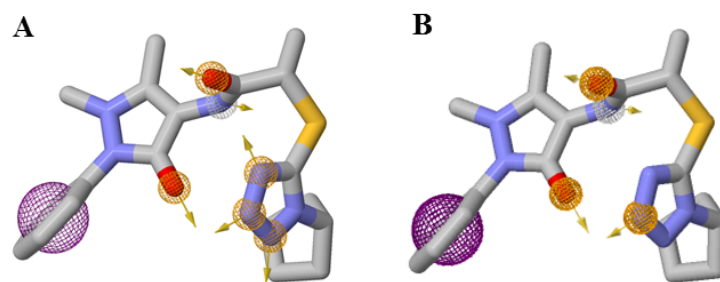


Figure 2.5 - representation of the two pharmacophore models built to find alternative molecules to ZINC12677264.

The docking method was optimised in such a way molecules that pose in a similar way to ZINC12677264 would be favoured. Therefore, constraints were introduced including hydrogen bond constraints and a region constraint. The hydrogen bond constraint definition included: Arg222 ( $\eta$ -NH<sub>2</sub>), Glu225 ( $\delta$ -CO<sub>2</sub>), Gln229 ( $\epsilon$ - $\eta$ -NH<sub>2</sub>), Gly252 (backbone CO) and Val254 (backbone NH and CO). The region constraint was based on an area of 3 Å around Phe218 and should privilege molecules that establish hydrophobic contacts with this region, specifically with aromatic ring. In terms of solutions, 100 best solutions were saved into the solutions file, since at this stage it was desirable to have a molecule that not only was positioned in a favourable way but also the pose was consistent enough to increase the levels of confidence about the chosen candidate. Compounds were docked using Goldscore and rescored using the remaining scoring options, similarly to what was done for the first set of molecules. To select the candidate molecule all molecules with a solubility value lower than -4 were removed, leaving only 29 candidate molecules. A multi-parameter analysis was then performed based on two ranks, ligand efficiency and RMSD, followed by a visual inspection. Unfortunately, none of the first ranked molecules seemed to pose in a way that would make possible the existence of aromatic  $\pi$ - $\pi$  stacking interaction with Phe218. In reality, only the sixth ranked molecule (ZINC63526256) seemed to present some docking poses consistent with this interaction. Alternatively, most of them seemed to interact with Tyr247 instead. The second best ranked molecule, ZINC64398049, was chosen based on having a very similar rank position to the first hit but having slightly better drug-like properties (*i.e.* number of hydrogen donors) and also due to the most consistent docking pose show the interaction with the hydrophobic pocket in addition to an aromatic  $\pi$ - $\pi$  stacking interaction with Tyr 247 – Figure 2.6.

Finally, three compounds were selected to screen using solution NMR -

ZINC08389070, ZINC71025726 and ZINC64398049. ZINC08389070, ZINC71025726 were obtained through a pharmacophore search based on the SxIP motif, whereas the last one, ZINC64398049, was obtained through a pharmacophore based on the molecule ZINC12677264. This second method has the advantage of providing larger chemical diversity, but the specificity given by the pharmacophore built on the natural ligand SxIP may be lost. To evaluate the robustness of the virtual screening and compound selection methods employed the *in vitro* screening of these molecules will be subsequently described.

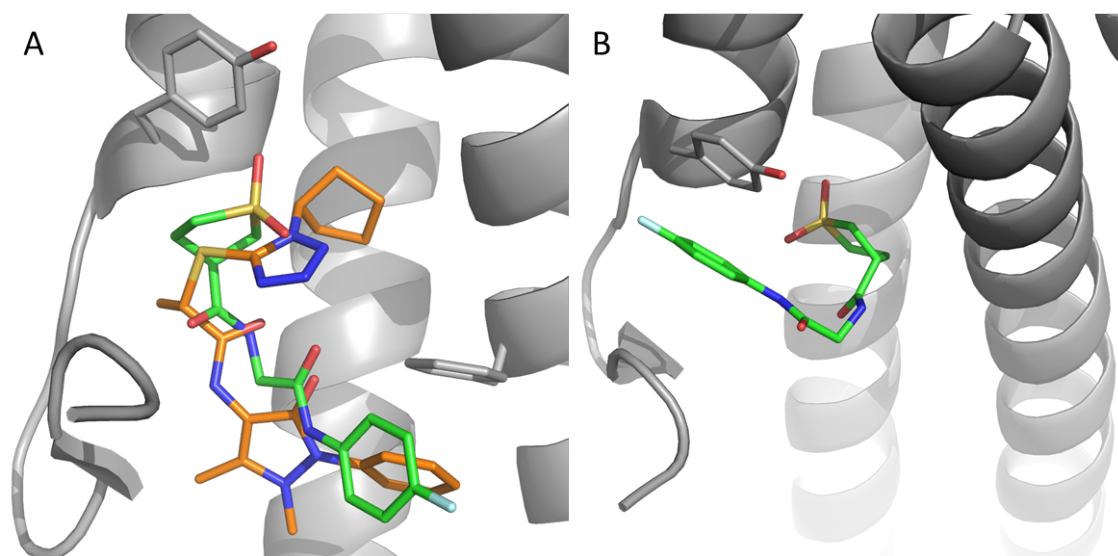
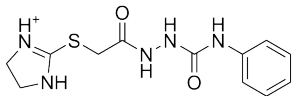
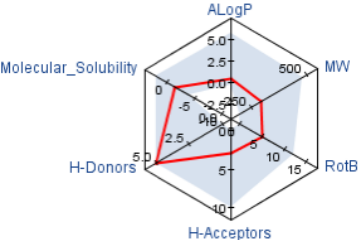
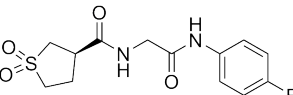
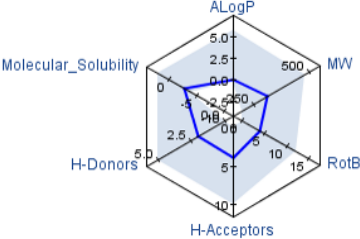
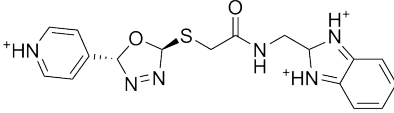
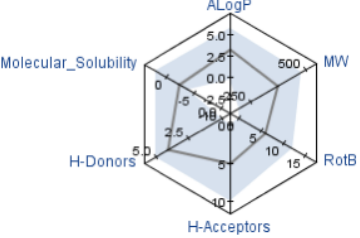


Figure 2.6 - ZIN64398049, A – ZINCPharmer result (coloured in green), based on ZINC12677264 (coloured in orange). B - Best scored pose. The aromatic ring is facing the ring of Tyr247 and is within a distance of 3-4 Å, suggesting the possibility of  $\pi$ - $\pi$  stacking interaction.

Identification of small molecule inhibitors

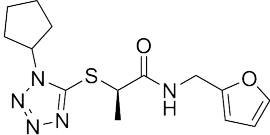
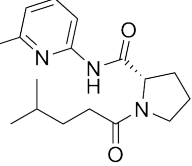
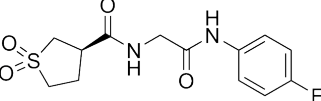
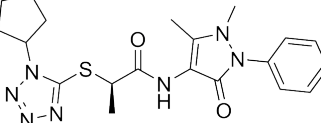
Table 2.9 - Summary of rankings and molecular properties for the three candidate molecules resulting from a pharmacophore model based on ZINC12677264. Final rank corresponds to the Pareto analysis using rank 1, score and ligand efficiency, and rank 2, RMSD between the obtained solutions.

Molecule Name	Final Rank	Rank 1	Rank 2	Solubility	Structure	Molecular Properties
ZINC44549656	1	1	2	-1.546		
ZINC64398049	2	1	3	-2.793		
ZINC63526256	6	3	4	-2.375		

## 2.6 NMR based screening for hit candidates based on SxIP motif

In the initial stage three compounds were screened for binding to EB1cΔ8 using  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC experiments – **1a**, corresponds to ZINC08389070, **1b**, based on ZINC71025726 synthesised in house and **1c**, ZINC64398049. Compound **1d**, ZINC12677264, was later purchased and tested due to the large similarity with compound **1a** – Table 2.10.

Table 2.10 - Summary of the candidate hit molecules resulting from virtual screening studies to be tested against EB1cΔ8

Code	ZINC code	Solubility	Molecular Weight	Gold-Score	Structure
<b>1a</b>	ZINC08389070	-3.96	321.4	55.9363	
<b>1b</b>	ZINC71025726	-3.785	289.4	55.3401	
<b>1c</b>	ZINC64398049	-2.793	314.3	52.4107	
<b>1d</b>	ZINC12677264	-3.992	427.5	62.3789	

The  $^1\text{H}$  and  $^{15}\text{N}$  resonances were fully assigned using complementary pairs of triple resonance NMR spectra - CBCA(CO)NH $^{13}\text{C}$ /HNCACB $^{137}\text{C}$  for  $\text{C}_\alpha/\text{C}_\beta$  and HNC $^{138-140}\text{C}$ /HN(CA)CO $^{141}\text{C}$  for CO connectivities. Backbone resonance assignment will be detailed in the next chapter – Chapter 3.

Ligand-induced chemical shift perturbations (CSPs) in NH resonances on addition of the ligand were used as an indication of ligand binding and location of the binding site. All compounds were shown to induce CSPs in the backbone of EB1cΔ8, with **1a** and

**1d** displaying the largest spectral changes - Figure 2.7, Figure 2.8, Figure 2.9, Figure 2.10 and Figure 2.11. No broadening was observed for the NH cross-peaks throughout the titration, indicating a fast exchange between the free and bound state and, therefore, weak interaction.

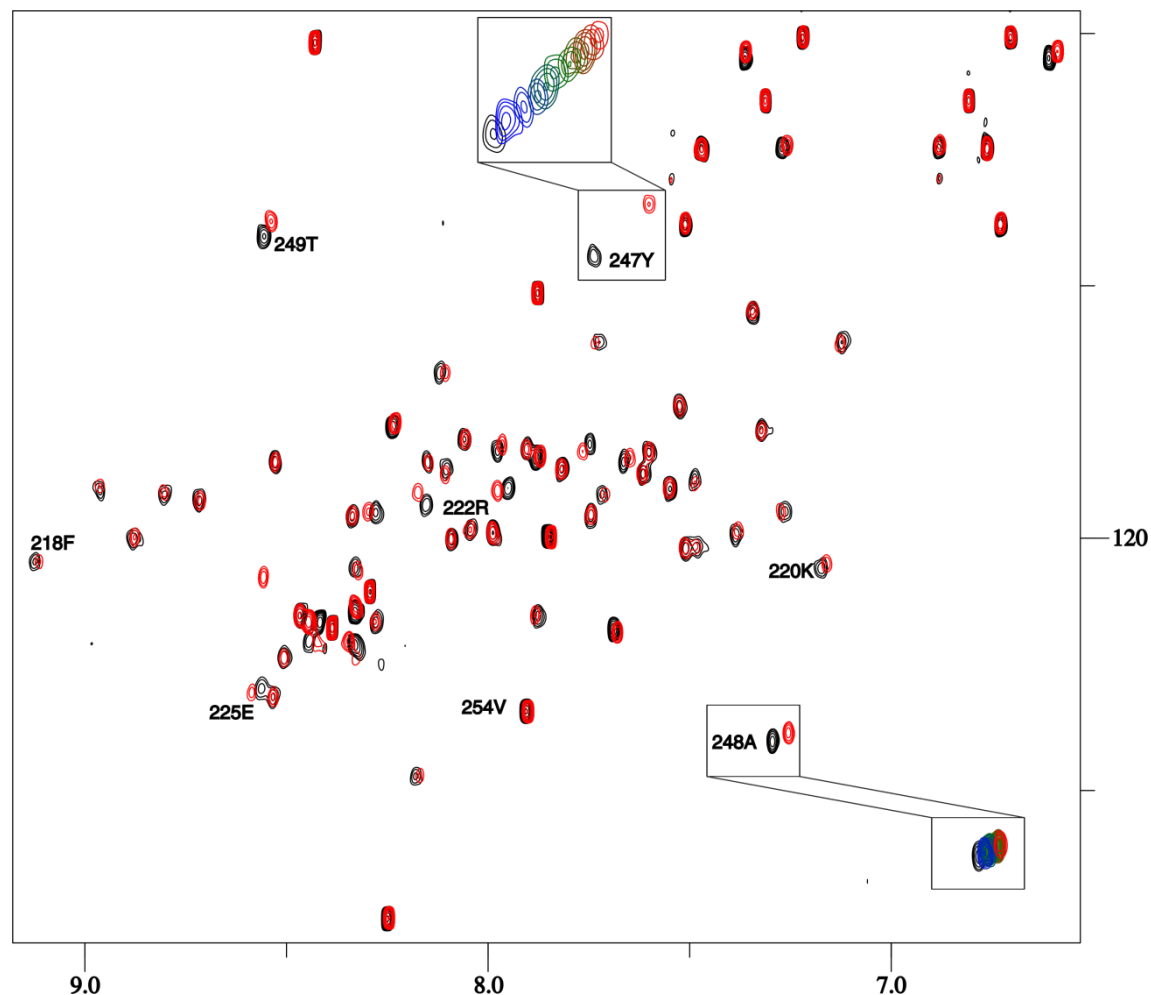


Figure 2.7 - Overlay of  $^1\text{H},^{15}\text{N}$ -HSQC spectra recorded at 800 MHz – in black is presented EB1c $\Delta$ 8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1c $\Delta$ 8 (50  $\mu\text{M}$ ) – **1a** (5000  $\mu\text{M}$ ). The insets show regions of interest and the spectra corresponding to the following **1a** concentrations, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 and 5000  $\mu\text{M}$ .

Compound **1a** shows two main regions where chemical shift changes are located -  $^{247}\text{YAT}^{249}$  and  $^{219}\text{GKLR}^{222}$ . Overall, the NH resonance for Tyr247 is the most affected upon ligand binding, with  $\Delta\delta = 0.42$  ppm - Figure 2.7.

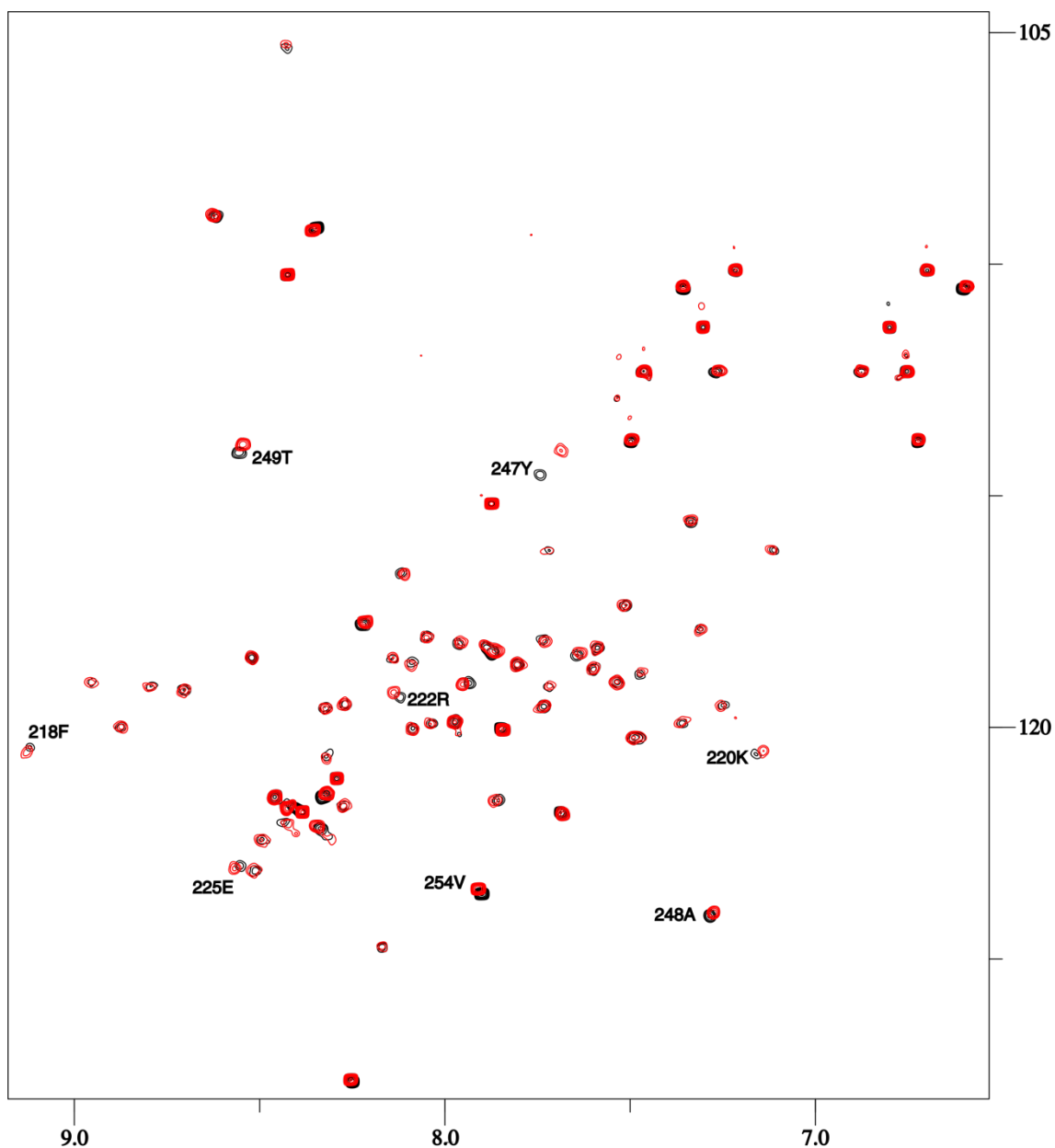


Figure 2.8 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1c $\Delta$ 8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1c $\Delta$ 8 (50  $\mu\text{M}$ ) – **1b** (5000  $\mu\text{M}$ ).

For compound **1b**, Tyr247 is again the most affected residue, followed by Thr249, Arg222 and Lys220 – Figure 2.8. All these CSPs are smaller than the ones observed for compound **1a**, indicating this is a weaker binder. It is worth to mention that all the CSPs are close to the hydrophobic pocket of EB1c $\Delta$ 8

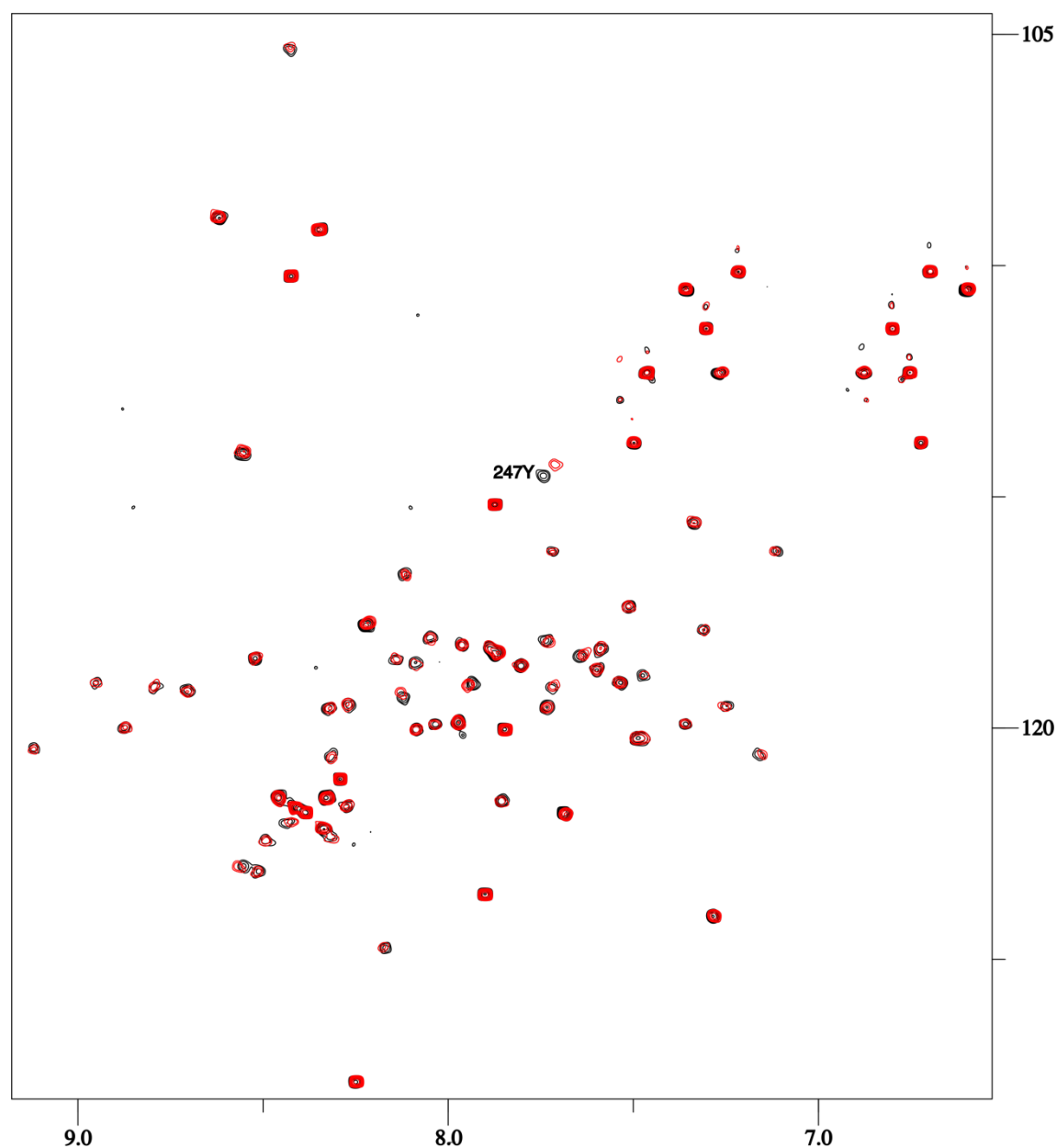


Figure 2.9 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1cΔ8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1cΔ8 (50  $\mu\text{M}$ ) – **1c** (5000  $\mu\text{M}$ ).

Compound **1c**, seems to be the worst ligand and only promotes CSPs for Tyr247 – Figure 2.9. Of the three initially tested compounds, **1a**, showed to be the most promising in terms of magnitude of CSPs and affected residues. Compound **1d** presents a similar scaffold to **1a**, and seems to perform better in terms of scoring in docking studies, especially the possible interaction of its aromatic ring with Phe218, that would significantly improve the potency of the molecule.



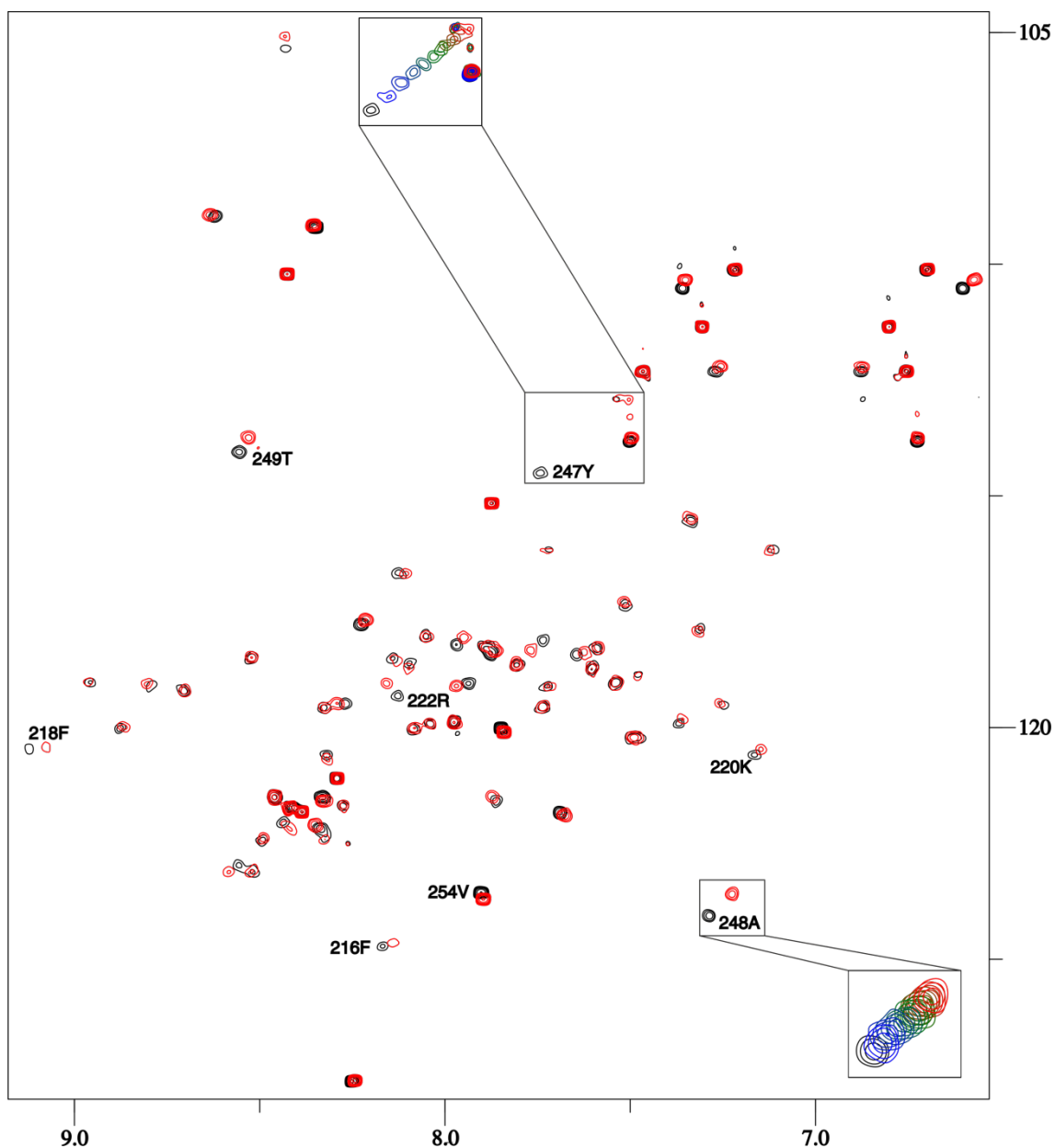


Figure 2.10 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1c $\Delta$ 8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1c $\Delta$ 8 (50  $\mu\text{M}$ ) – **1d** (5000  $\mu\text{M}$ ). The insets show regions of interest and the spectra corresponding to the following **1d** concentrations, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 and 5000  $\mu\text{M}$ .

$^1\text{H}$ ,  $^{15}\text{N}$ -HSQC titration confirms partially this hypothesis. Tyr247 is, once again, the most affected residue, the CSPs still follows a fast exchange pattern, but the distance is larger when compared with **1a** ( $\Delta\delta = 0.66$  ppm compared with 0.42 ppm obtained for **1a**). Interestingly,  $^{247}\text{YAT}^{249}$  and  $^{219}\text{GKLR}^{222}$  are still the main regions where chemical shift changes are located. Additionally, Phe218, Phe216 and even Val254 ( $\Delta\delta$  is rather small, 0.06 ppm) are perturbed (Figure 2.10), indicating this compound interacts not only with the hydrophobic pocket usually occupied by the IP motif of

SxIP proteins but also with the aromatic rim just below it and in a smaller extent with the dynamic C-terminus. It is worth mentioning that Val254 seems to be an important residue for the interaction of EB1cΔ8 with other molecules, being involved in hydrogen bond contacts with MACFp1 as described by Honnapa et al.<sup>31</sup> In Figure 2.11 it is clear that compound **1d** promotes the largest chemical shift changes followed by compound **1a**.

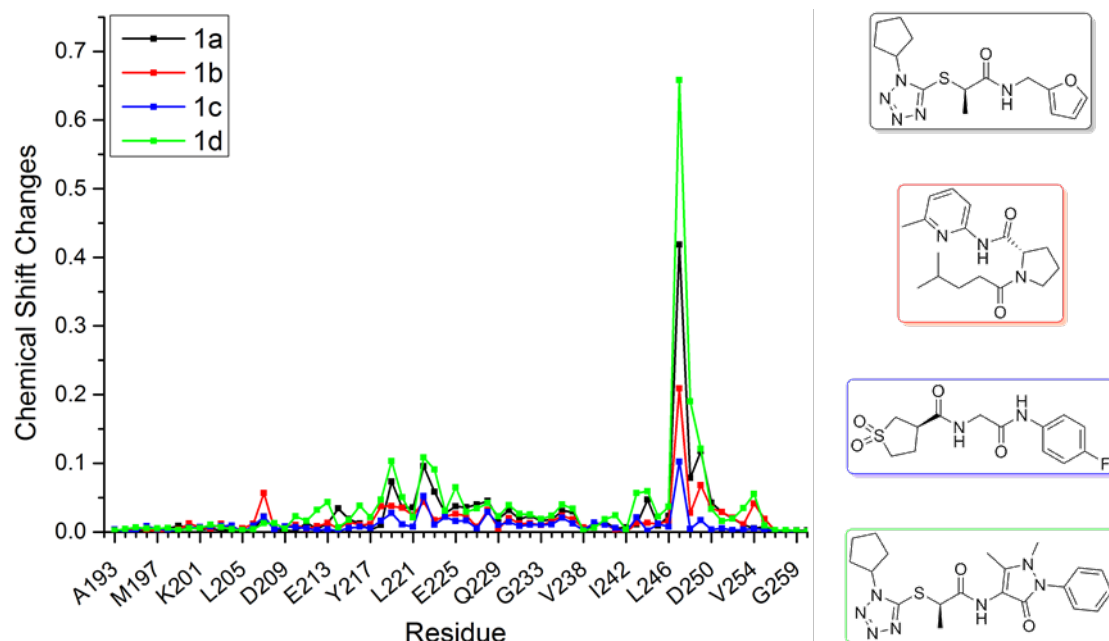


Figure 2.11 - Chemical shift changes plot for the four tested compounds and their distribution *per* EB1cΔ8 residue.

NMR titration curves show significant deviation from a linear dependence at high ligand excess that allows for  $K_d$  estimation by fitting the curve into the two-state exchange model (Figure 2.12). In agreement with the CSP amplitudes, the estimated binding affinity is higher for **1d** ( $6 \pm 1$  mM) followed by compound **1a** ( $10 \pm 3$  mM) - Table 2.11.

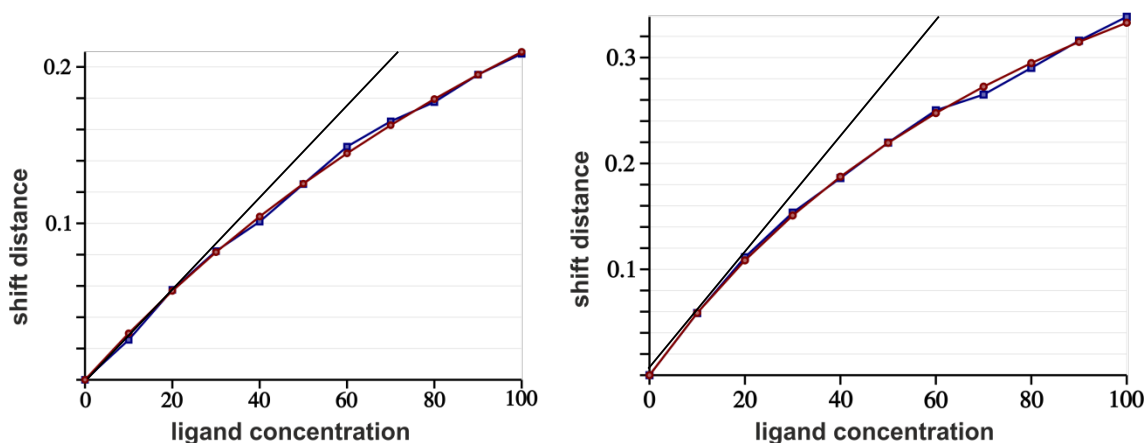


Figure 2.12 - Chemical shift changes plots for Tyr 247, upon titration with **1a** (left hand side plot) and **1d** (right hand side panel).

Table 2.11 – values used to estimate  $K_d$  values for interaction between EB1c $\Delta$ 8-**1a** and **1d**.

<b>1a</b>		<b>1d</b>	
Residue	$K_d$ (mM)	Residue	$K_d$ (mM)
Thr206	6.05	Tyr247	5.29
Asp209	7.22	Ala248	7.37
Asp2015	15		6.33 ( $\pm$ 1.04)
Arg241	10		
Tyr247	10		
9.65 ( $\pm$ 3.09)			

Both best binding compounds **1a** and **1d** share the same scaffold (Figure 2.13). This scaffold has shown to have high specificity (chemical shift mapping and molecular docking predictions) to the IP binding site. Moreover, these are the first molecules reported to bind to EB1c indicating that the aims of the chapter were achieved. One can argue that the binding affinities are not very high (mM); however, it is very close to the one showed by the SxIP motif - Figure 2.14. The curvature is more pronounced for compound **1d**, followed by **1a** and finally SKIP peptide, indicating the latter is closer to linearity and therefore weaker binding. Full information on the titration of this peptide can be found in Chapter 5. Analysis of the docking poses obtained for these compounds during the virtual screening and comparison with the IP motif of MACFp1 in the crystal structure of the complex indicates that this scaffold may act as an IP motif mimetic. The hydrophobic side chain of the isoleucine is replaced by a cyclopentyl ring, and the hydrophobic proline ring is replaced by a methyl group - Figure 2.13. The tetrazole moiety acts as spacer between both hydrophobic regions giving rigidifying the scaffold.

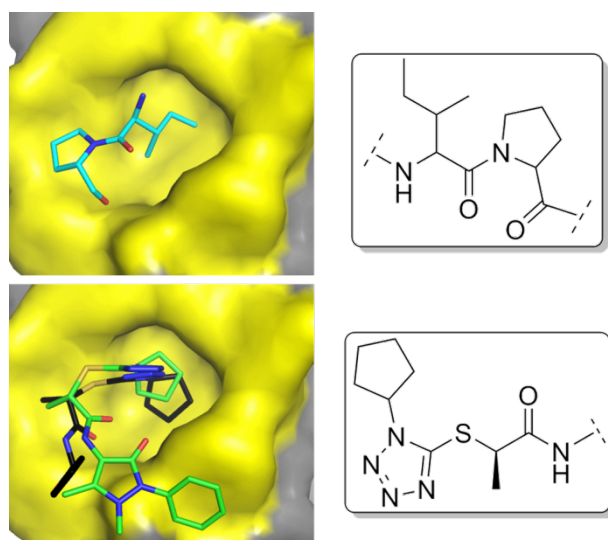


Figure 2.13 – Left panel - 3D model for the IP motif and the IP motif mimetic. The IP motif tri-dimensional representation is based on the crystal structure 3GJO. The IP mimetic compounds **1a** (black) and **1d** (green) are the binding poses predicted by our docking studies using 3GJO structure as the EB1cΔ8 model. In both representations the C-terminus tail was removed for clarity. Right panel - 2D structure of IP motif and IP mimetic scaffold.

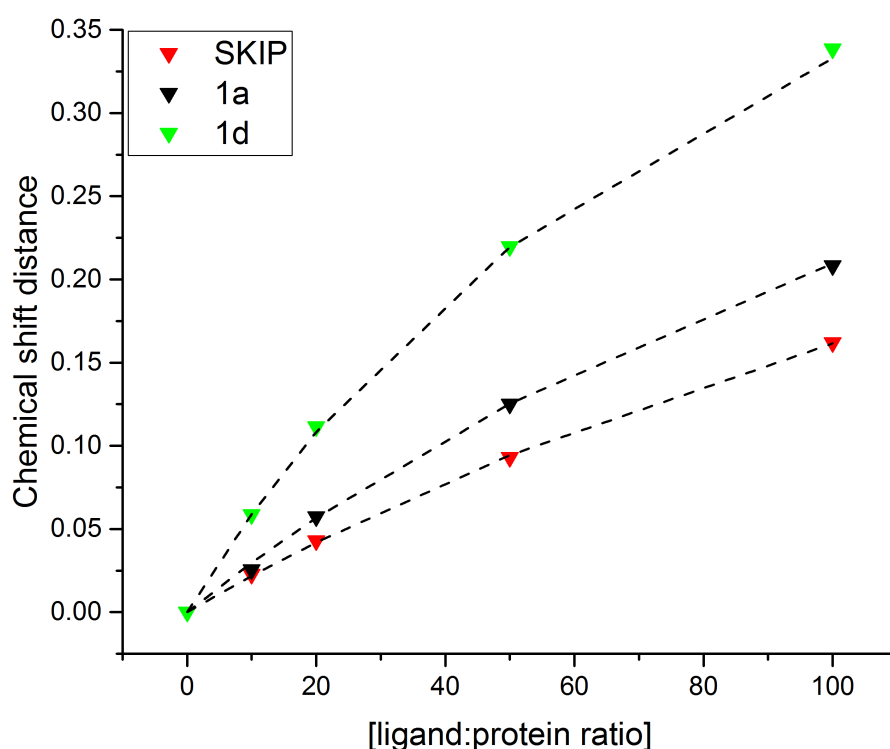


Figure 2.14 – Binding curve for residue Tyr247 for the titrations of the tetramer **SKIP** (red), compound **1a** (black) and compound **1d** (green). Ligand concentration corresponds to the excess of ligand (*e.g.* 10 fold excess, 20 fold excess) to the protein concentration (50  $\mu$ M). The fitted curve performed through the equation  $y=A((B+4x-\sqrt{(B+4x)^2-4x})$  is shown as black dashed line.

It is therefore very likely that this scaffold is very specific for this binding site. To validate the docking predictions and to gain a structural insight of the binding mode of the IP mimetic scaffold the three dimensional NMR structure of EB1c $\Delta$ 8 bound to compound **1a** was determined. Additionally, the structure of EB1c $\Delta$ 8 in the free state was also elucidated in order to identify structural changes that may occur upon compound binding and to facilitate structure determination of the complex. These structures and other aspects considering the binding mode of compound **1a** to EB1c $\Delta$ 8 will be the subject of the next chapter.

## Chapter 3. STRUCTURAL CHARACTERISATION OF EB1C USING SOLUTION NMR

It was reported, in the previous chapter, the identification of a IP mimetic scaffold. It is now fundamental to understand how this scaffold interacts with EB1 and gain structural information on this complex to help to develop higher affinity ligands. For that purpose, EB1c will be characterised in the free form and in complex with compound **1a**. Comparison between the two forms and analysis of the binding mode of this scaffold will be also analysed.

### 3.1 Design of EB1 fragments

It has been reported the importance of C-terminal tail on EB1 binding. Therefore, and to satisfy the aims of the project it was decided to build two different gene constructs to express two versions of the C-terminal of EB1 (EB1c). EB1c $\Delta$ 8 construct comprises residues 191-260 and EB1c $\Delta$ 16 corresponds to residues 191-252. The choice of these two different constructs was based on the necessity of a better understanding of the role of flexible tail for the interaction. Full-length EB1c (191-268) was not used in this project as the literature indicate the EB1c $\Delta$ 8 includes all the necessary amino acids for the interaction with SxIP proteins.

### 3.2 Production of recombinant EB1c $\Delta$ 8 and EB1c $\Delta$ 16

The necessary steps to obtain the two EB1c fragments – EB1c $\Delta$ 8 (residues 191-260) and EB1c $\Delta$ 16 (residues 191-252), such as cloning, transformation, expression and purification are described in more detail in Chapter 7. EB1c fragments were expressed in an N-terminal SUMO vector (pOPINS) and high levels of protein expression were achieved. The protein was subsequently purified with good yields – Figure 3.1. The purification process included Ni affinity chromatography, followed by removal of the histidine tag with the use of SUMO protease. This particular step required optimisation because the protease was not cleaving the tag with complete efficiency, leaving some SUMO-tagged EB1c. This problem was overcome by using a reducing agent, 2-mercaptoethanol, performing the cleavage at room temperature and using higher

concentrations of the protease. Finally, ion exchange purification was performed. The SDS-PAGE gels in Figure 3.1 show the steps performed until pure protein was obtained. The first corresponds to the purification of EB1c $\Delta$ 8 and the lanes represent the following: 1 – Ni<sup>2+</sup> affinity column flow through, 2 - Ni<sup>2+</sup> elution with high imidazole buffer, 3 – buffer exchange into no imidazole buffer, 4 – after SUMO protease cleavage, 5 – Ni<sup>2+</sup> reverse purification, 6 – SUMO tag, 7 – pure protein after ion exchange chromatography. The second gel corresponds to EB1c $\Delta$ 16 and the lanes follow the subsequent pattern: 1 – before Ni<sup>2+</sup> affinity chromatography, 2 – Ni<sup>2+</sup> affinity column flow through, 3 - Ni<sup>2+</sup> elution with high imidazole buffer, 4 – after SUMO protease cleavage, 5 – Ni<sup>2+</sup> reverse purification, 6 – SUMO tag, 7 – pure protein after ion exchange chromatography. The chromatograms obtained for Ni<sup>2+</sup> affinity chromatography and ion exchange chromatography are shown in Figure 3.2.

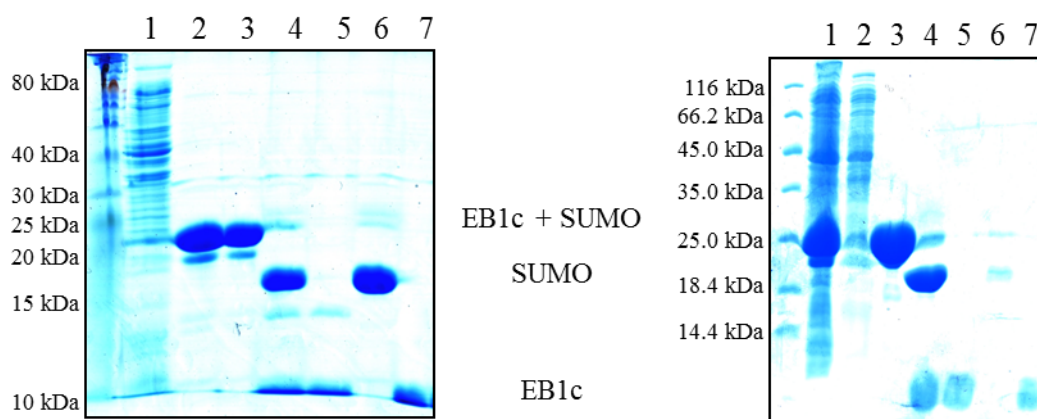


Figure 3.1- SDS-PAGE electrophoresis of purified EB1 fragments using 15% Tris-Glycine gels – EB1c $\Delta$ 8 (8 kDa) and EB1c $\Delta$ 16 (7.2 kDa). Molecular weights of the proteins expressed were predicted using ProtParam webserver<sup>14</sup>.

Unlabelled, <sup>15</sup>N-labelled, <sup>13</sup>C<sup>15</sup>N-labelled EB1c $\Delta$ 8, as well as unlabelled and <sup>15</sup>N-labelled EB1c $\Delta$ 16 samples were characterised using a range of NMR spectroscopic experiments. <sup>1</sup>H 1-D spectra suggest that both fragments are folded in solution, showing an overall dispersion of chemical shifts in the methyl (0-2 ppm) and amide (8.5-10 ppm) regions - Figure 3.3 and Figure 3.4.

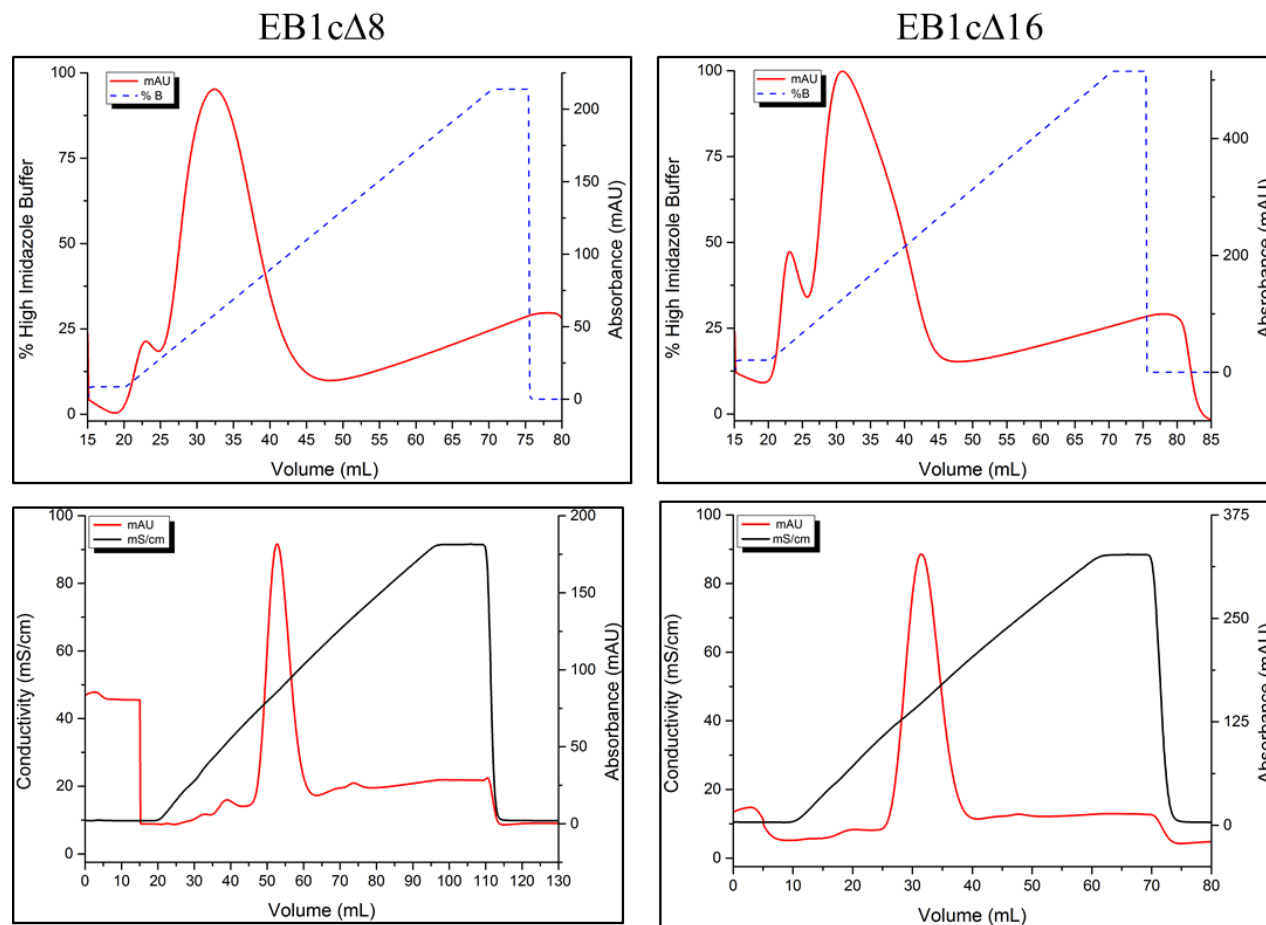


Figure 3.2 - Chromatograms for EB1cΔ8 and EB1cΔ16 purifications. Upper panel – Chromatogram obtained for Ni affinity chromatography for EB1cΔ8 (left hand side) and EB1cΔ16 (right hand side). Bottom panel – Chromatogram obtained for ion exchange chromatography for EB1cΔ8 (left hand side) and EB1cΔ16 (right hand side). Both proteins were used without further purification after ion exchange chromatography.



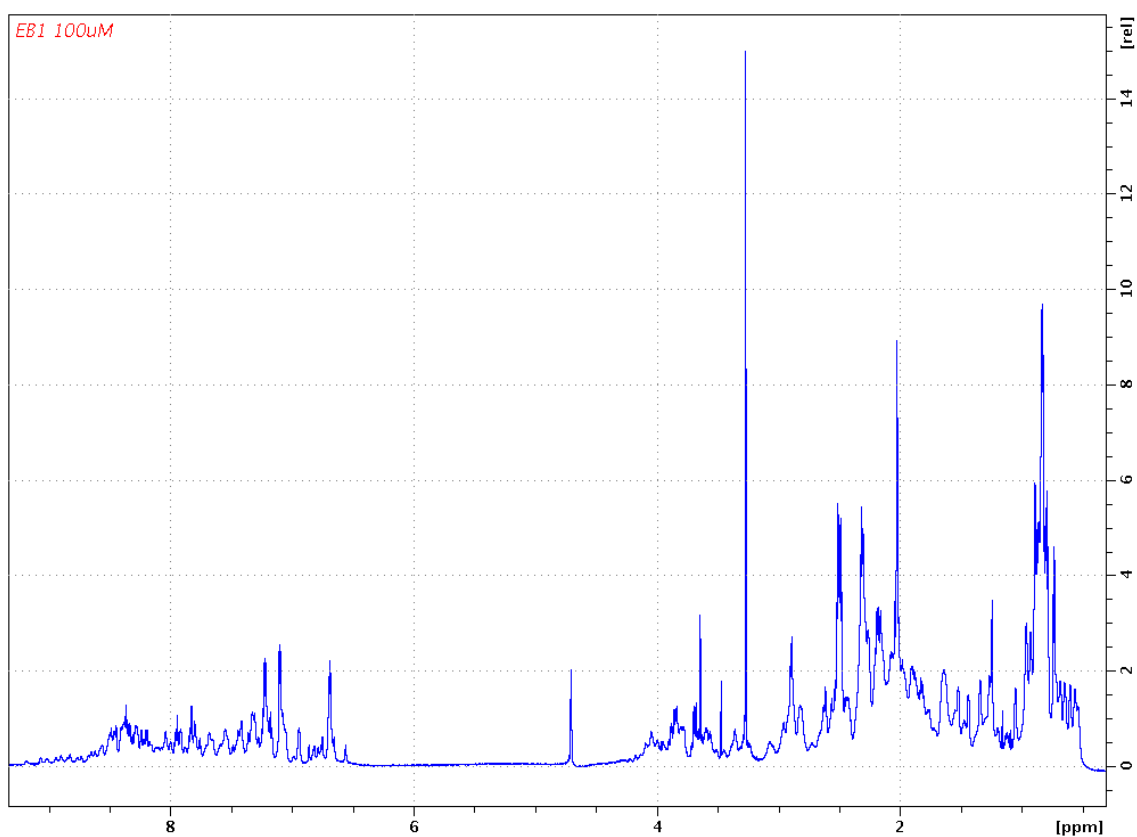


Figure 3.3-  $^1\text{H}$  1-D spectra of  $^{15}\text{N}$ -EB1c $\Delta$ 8 at 100  $\mu\text{M}$  acquired at 25°C, with a field strength of 600 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02%  $\text{NaN}_3$ .

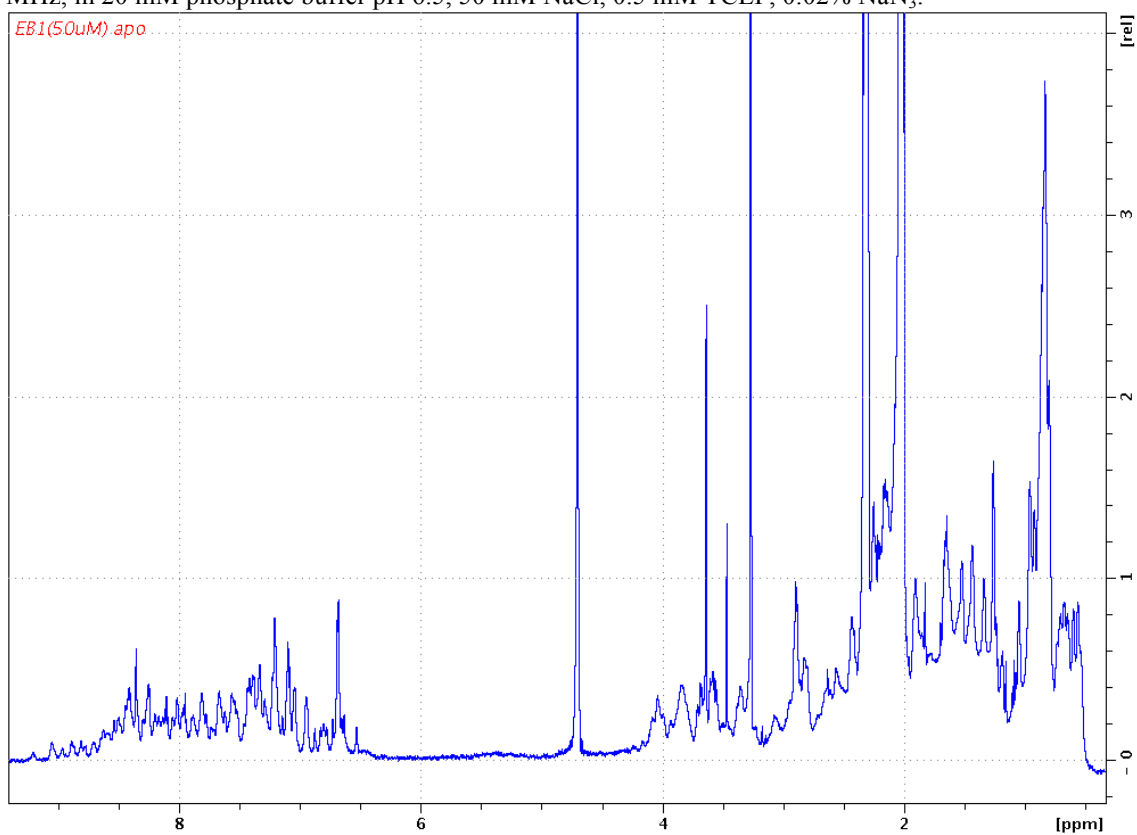


Figure 3.4 -  $^1\text{H}$  1-D spectra of EB1c $\Delta$ 16 at 50  $\mu\text{M}$  acquired at 25°C, with a field strength of 600 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02%  $\text{NaN}_3$ .

### 3.3 Characterisation of EB1c by NMR Spectroscopy

The C-terminal domain of EB1 has been previously structurally characterised and described as a stable dimer with a parallel coiled coil<sup>31</sup> Each monomer starts with a long smoothly curved helix (residues 191-230), followed by a loop region which connects to a shorter second helix (237-248) antiparallel to the longer helix, forming a parallel homodimer characterised by the existence of a coiled coil region followed by a four helix bundle. The four helix bundle is then followed by a highly disordered and flexible C-terminal region.<sup>27,31,43</sup> Our construct EB1cΔ8 includes most of the C-terminal flexible region since this region seems to establish contacts with ligands.<sup>31</sup>

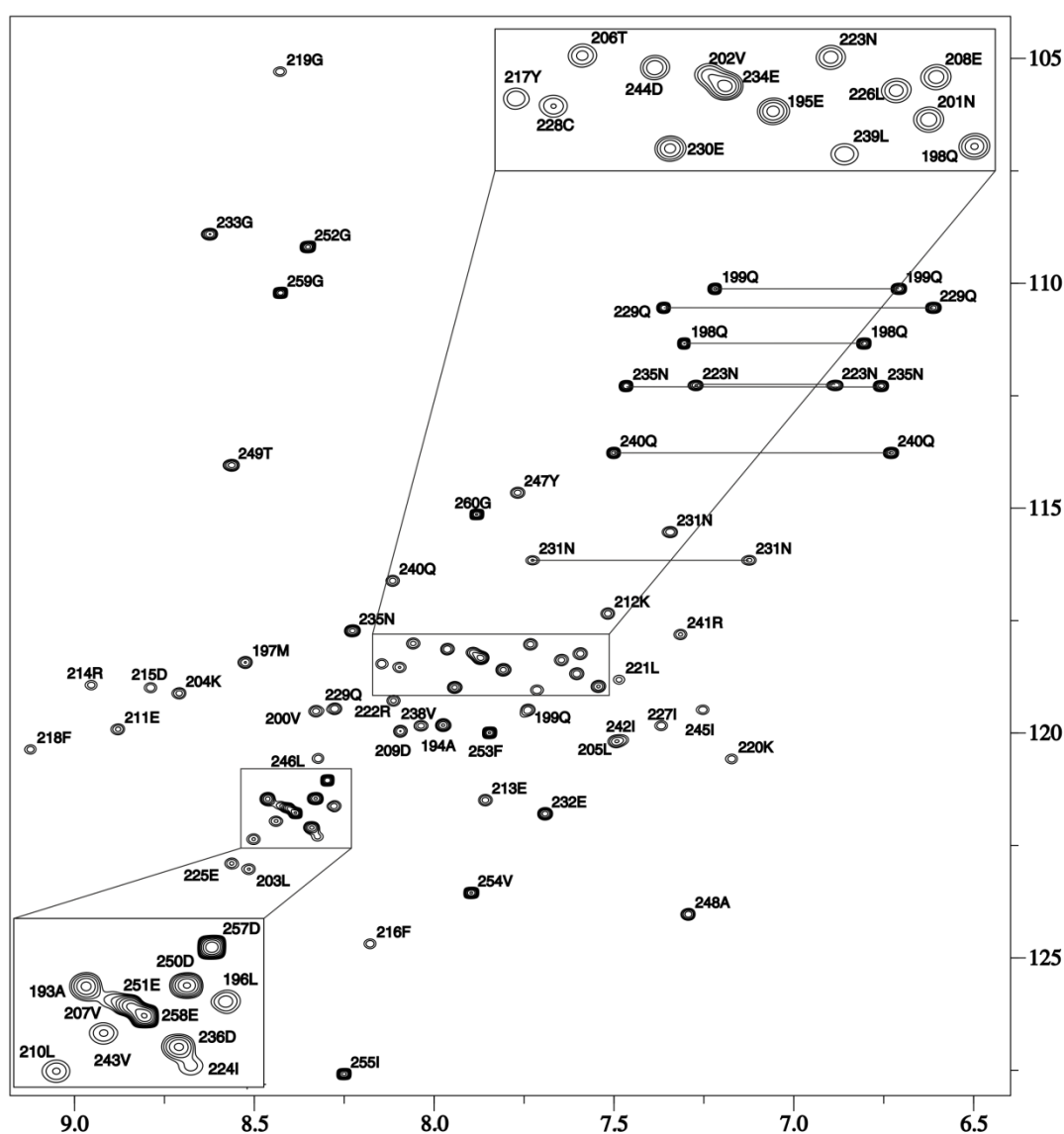


Figure 3.5 -  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum for EB1cΔ8 acquired at 25°C, with a field strength of 800 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02%  $\text{NaN}_3$ . The spectrum shows the protein is folded and the well dispersed peaks allowed a complete backbone assignment.

Both  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra (Figure 3.5 and Figure 3.6) present large chemical shift dispersion for the NH resonances. The good dispersion is related to protein folding meaning that the secondary structure places the residues in defined regions, making each residue being in a unique position that will subsequently yield unique NMR resonance signals. The main difference is the existence of extra peaks in the EB1c $\Delta$ 8's spectrum as expected as this construct is larger.

The  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC of EB1c $\Delta$ 8 was fully assigned by using a sequential assignment methodology with complementary pairs of NMR spectra - CBCA(CO)NH/HNCACB for  $\text{C}\alpha/\text{C}\beta$  and HNC(O)/HN(CA)CO for CO assignment (Figure 3.7). The use of these spectra is based on their C correlation with the NH resonances. The CBCA(CO)NH/HNCACB relation of  $i/i-1$  was used for the sequential assignment building residue connections. Additionally, by estimating the residue type using  $\text{C}\alpha/\text{C}\beta$  chemical shifts to it was possible to estimate possible sequences and integrate them in EB1c $\Delta$ 8 in an iterative way, assigning larger residues stretches with the progress of the assignment.

The HNC(O)/HN(CA)CO  $i/i-1$  pair, was used to resolve possible ambiguities. In Figure 3.7 one can observe the overall quality of the CBCA(CO)NH/HNCACB pair is good, with good dispersion and resolution of cross-peaks. However, the same region in the HNC(O)/HN(CA)CO pair has a much better quality, showing sharper peaks resolved and isolated from neighbouring peaks. It is important to mention that the CO chemical shifts can also be used to help predict secondary structure of the protein. The assignment procedures are described in Chapter 7.

With the backbone assignment of EB1c $\Delta$ 8, all NH resonances present in the  $^1\text{H}$ - $^{15}\text{N}$ -HSQC were assigned with the exception for the side chains. The NH resonances for the other construct were easily assigned by superimposition of this spectrum on the previously assigned EB1c $\Delta$ 8  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC. The C-terminal residues for EB1c $\Delta$ 8 are, as expected, more intense – Figure 3.5, left hand side inset, residues Asp257 and Glu258 indicating this region is highly dynamic.

Alternatively, the spectrum for the construct EB1c $\Delta$ 16 only shows high intensity peaks for Asp250 and Glu251. As described before this construct lacks 16 C-terminal residues and therefore lacks the C-terminal flexible tail. The main purpose of this construct, EB1c $\Delta$ 16, was to assess the role of the C-terminal flexible region – last eight

residues, in the interaction with other molecules. Therefore, the characterisation of this construct will not be as thorough (side chain assignment) as the one performed for EB1c $\Delta$ 8 and subsequently presented.

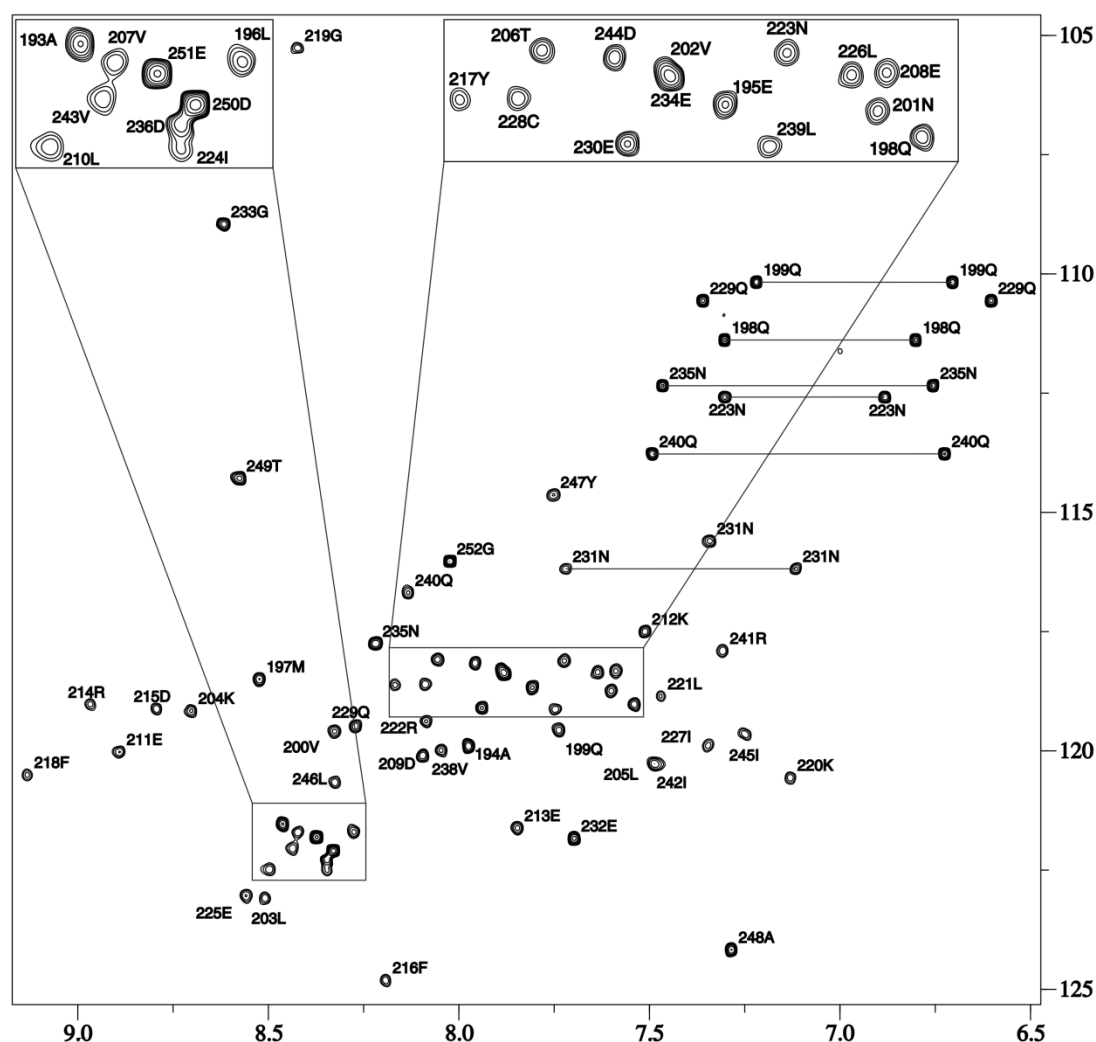


Figure 3.6 -  $^1\text{H}, ^{15}\text{N}$  – HSQC spectrum for EB1c $\Delta$ 16 acquired at 25°C, with a field strength of 600 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02%  $\text{NaN}_3$ . The spectrum shows the protein is folded and the well dispersed peaks allowed a complete backbone assignment.

Using the HBHA(CO)NH experiment that correlates NH of  $i$  to  $\text{H}\alpha/\text{H}\beta$  of  $i-1$  it is possible to assign  $\text{H}\alpha/\text{H}\beta$  resonances from the NH resonances. The assignment of this spectrum was also a confirmation of the NH resonances assignments since  $\text{H}\alpha/\text{H}\beta$  resonances also have characteristic values for each amino acid type and all  $i/i-1$  links are confirmed since this experiment works on  $i/i-1$  relation.<sup>143–144</sup>>

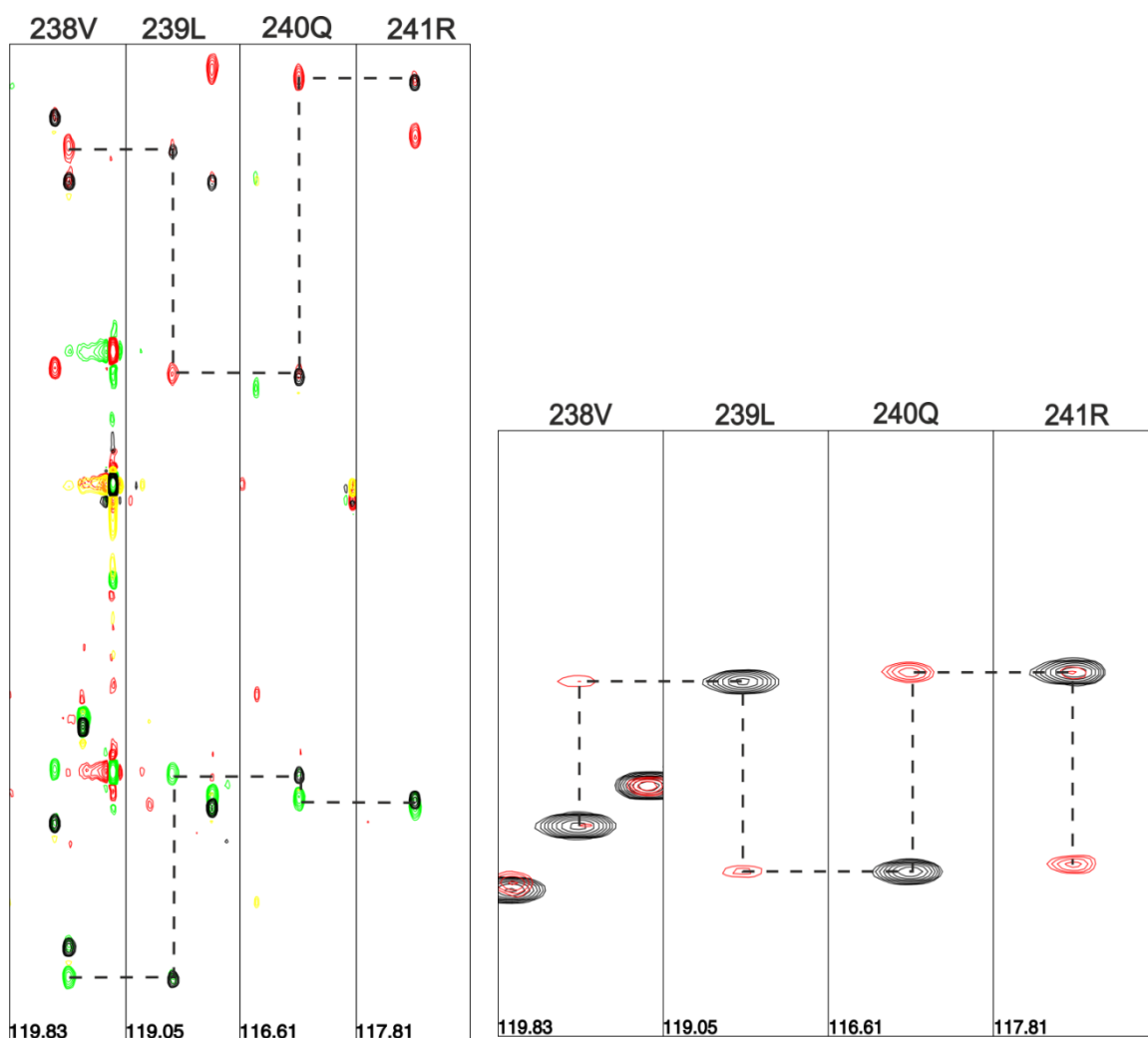


Figure 3.7 – Superimposition of the pairs - CBCA(CO)NH/HNCACB and HNC(OH)(CA)CO. Left panel - the  $C_{\alpha}/C_{\beta}$  peaks of the CBCA(CO)NH spectrum are shown in black, with the  $C_{\alpha}/C_{\beta}$  peaks of the HNCACB spectrum are shown in green and red respectively. Right panel – the HN(CA)CO spectrum is shown in red, with the  $C'$  peak of the preceding residue being less intense and matching the  $i-1$  peak from the HNC(O) spectrum.

In general, it is not sufficient to obtain only the backbone assignments, since side chains are the major fraction of the protein. Moreover, no protein structure can be determined without side chain assignments.

Hence,  $^1\text{H}, ^{13}\text{C}$ -HSQC (Figure 3.8) was assigned, starting with the known resonances for  $\alpha/\beta$  resonances, previously obtained from the combination of CBCA(CO)NH/HNCACB and HBHA(CO)NH. Using these resonances and the through bond correlations given by H(C)CH-TOCSY experiment the remaining atoms in the side chain were assigned. Because the resolution of  $^1\text{H}, ^{13}\text{C}$ -HSQC is inferior to the  $^1\text{H}, ^{15}\text{N}$ -HSQC and there are more  $^1\text{H}/^{13}\text{C}$  cross-peaks, the spectrum looks more crowded, with more overlapping, therefore, (H)CCH-TOCSY experiment was used to discriminate between signals with similar  $^1\text{H}$  resonances.

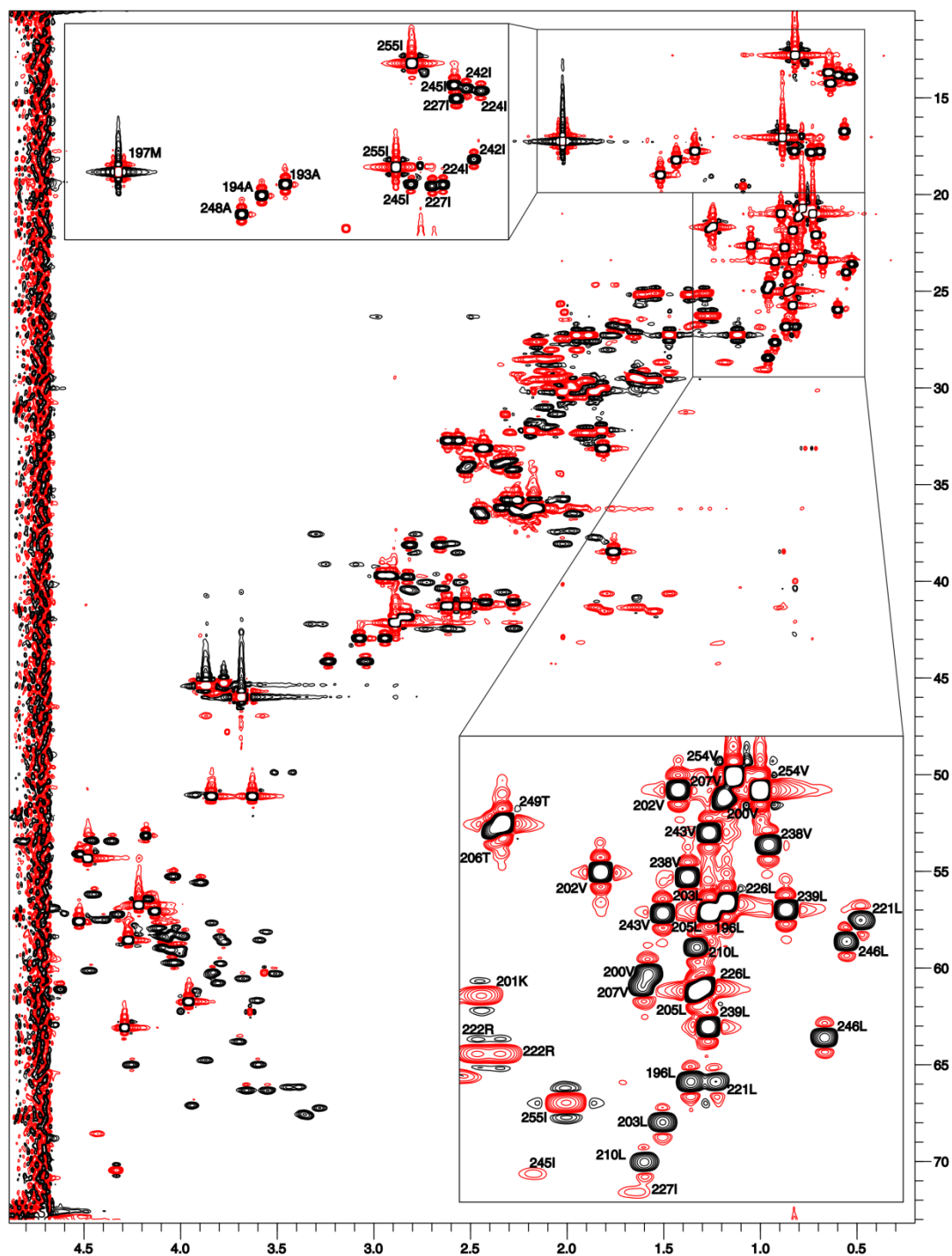


Figure 3.8 -  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC spectrum for EB1c $\Delta$ 8 acquired at 25°C, with a field strength of 800 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% NaN<sub>3</sub>. The spectrum shows generally well dispersed peaks. Only methyl region assignments are shown for clarity. Positive contours are shown in black and negative contours in red.

These experiments make use of the TOCSY (Total Correlation Spectroscopy) through-bond coherence transfer to all coupled spins in a scalar-coupled network (e.g. all  $^{13}\text{C}$  or  $^1\text{H}$  atoms of a linear side-chain).

The methyl region of  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC is very crowded, especially methyl groups of valines and leucines. The region around 25-35 ppm in carbon dimension and 1-2.3 ppm for hydrogen dimension is also very crowded and it was particularly difficult to assign. This region contains high density of peaks including  $\beta$ 's for a large number of residues – lysine, arginine, glutamic acid, glutamine, methionine, proline;  $\gamma$ 's for isoleucine, lysine, arginine, proline and  $\delta$ 's for lysine.

The aromatic residues resonances (Figure 3.9) were assigned using the  $\text{C}\alpha/\text{C}\beta$  and  $\text{H}\alpha/\text{H}\beta$  resonances and a combination of 2D and 3D NOESY spectra for the side chain assignment. The main reason for this approach was the poor quality of  $\text{H}(\text{C})\text{CH}$ -TOCSY spectra for the aromatic region. The signals of the aromatic side chains are not well dispersed and required a very careful analysis.

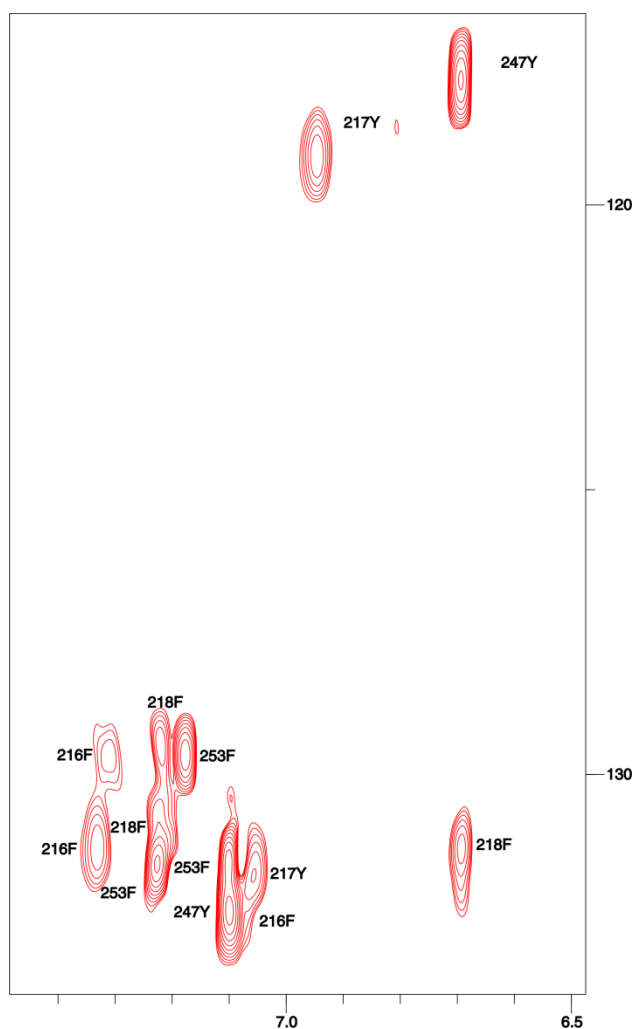


Figure 3.9 -  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC spectrum for the aromatic residues of EB1c $\Delta$ 8 acquired at 25°C, with a field strength of 800 MHz, in 20 mM phosphate buffer pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02%  $\text{NaN}_3$ .

Once side chain assignments were performed, the NOE spectra –  $^{15}\text{N}$ -resolved- $^{13}\text{C}$ -

decoupled-NOESY-HSQC and  $^{13}\text{C}$ -resolved-3D-NOESY-HSQC – HSQC dimensions were assigned using the root resonances present in both  $^1\text{H},^{15}\text{N}$ -HSQC and  $^1\text{H},^{13}\text{C}$ -HSQC, in a semi-automatic way, *i.e.* picking the peaks and propagating the assignments.

An extra complication of the EB1c resonance assignment is the fact that it is a symmetric parallel homodimer, making difficult to distinguish between intraresidue NOEs and intermonomer NOEs between the same residue belonging to different monomers. In other words, with the assignment of  $^{15}\text{N}$ -resolved- $^{13}\text{C}$ -decoupled-NOESY-HSQC and  $^{13}\text{C}$ -resolved-3D-NOESY-HSQC one is assigning NOE cross-peaks for only one monomer. However, due to the short distance between both monomers it is possible that some of these peaks, especially in the methyl region, are intermonomer contacts. It was, therefore, of the utmost importance to acquire information about NOE distances between the two monomers and therefore position the monomers in respect to each other – intermonomer distance restraints. For that purpose, EB1c $\Delta$ 8 and  $^{15}\text{N}^{13}\text{C}$ -EB1c $\Delta$ 8 equimolar solutions were mixed and incubated at 37°C for 16 hours. This procedure was previously described by De Groot et al<sup>14</sup> for an EB1/EB3 heterodimer formation. A set of  $^{15}\text{N},^{13}\text{C}$ -filtered NMR NOE experiments were acquired permitting the calculation of distance restraints between residues located in separate monomers (isotopically labelled *versus* unlabelled).

With the  $^{15}\text{N}$  and  $^{13}\text{C}$  resolved 3D-NOESY-HSQC root resonances fully assigned it is possible to introduce these chemical shifts and spectra into ARIA<sup>96</sup> for an automated ambiguous automated assignment. ARIA uses an iterative protocol and the concept of ambiguous distance restraints and automatically assigns NOE cross peaks.<sup>96</sup> In addition to NOE cross peaks and calculated interatomic distances the use of other restraints such as dihedral angle or hydrogen bond restraints can be very useful for NMR structure calculation. This will be the focus of the next section.

### 3.4 Solution NMR structure of EB1c $\Delta$ 8

ARIA<sup>96</sup> is a widely used automated NOESY assignment procedure that acts as an interface for the Crystallography and NMR system - CNS<sup>98</sup>. A very important concept introduced by ARIA is the use of ambiguous distance constraints, where a certain chemical shift can be assigned to more than one resonance and later resolved as the



calculation proceeds. The use of only unambiguously assigned NOEs hinders the automatic NOE assignment since the majority of NOEs are hard to assign unambiguously.<sup>97</sup> More information on ARIA 2.0 can be found in NOE Assignment with ARIA 2.0.<sup>14></sup>

The EB1c $\Delta$ 8 structure in the unbound state was determined from 2641 restraints from which 634 correspond to intermonomer distance restraints, supplemented by 64 dihedral angle restraints derived from chemical <sup>13</sup>C-chemical shift values. Statistics of the structure determination are presented in Table 3.1. The process is described in Chapter 7. From the 320 structures calculated for the last iteration, 30 were water refined. The overlay of the 20 lowest energy structures – Figure 3.10 – shows consistency across the 20 structures for the region 191-248 where a helical secondary structure can be observed. EB1c is a parallel 2-stranded coiled coil, where the side chain of an apolar residue from one chain is inserted into a hole formed by the apolar side chains of four residues from the opposing chain – “knob in the hole” like structure.<sup>14></sup> The coiled coil is then followed by a four helix bundle where contacts between four helices can be observed. Finally, a disordered C-terminus region follows, where assignments need to be carefully and manually analysed since any wrong or highly ambiguous assignment can introduce a bias in this region yielding an incorrect structure model.

The first difficulty encountered when trying to calculate an ensemble of structures using dihedral angles and distance restraints was, due to being a large continuous helical structure and to the ambiguity given by the large number of leucines and valines (11.4% and 8.6% of the total sequence respectively), the calculated monomer would invariably fold on itself. This structure yielded high number of violations and high energies being, in all aspects, an unfavourable conformation. However, there were not enough restraints to fold it without breaking the helix in the region between the leucine zipper and the four helical bundle structure (Leu210-Arg214). To overcome this problem it was thought that the introduction of hydrogen bond restraints could help to stabilise the helical structure of the protein. Therefore, a thorough analysis of the NOEs was carried out in order to confirm which regions of the protein have a helical conformation. A survey of the sequential and medium range NOEs states that for an  $\alpha$ -helix secondary structure, NOE cross peaks for H $\alpha$  of  $i$  and HN of  $i+2$  and  $i+3$  should be observed, as well as H $\alpha$  for  $i$  and H $\beta$  of  $i+3$ .<sup>15></sup>

Table 3.1 – NMR distance and dihedral statistics and structure statistics determined for the 20 best structures of the ensemble of free EB1cΔ8.

EB1cΔ8	
Total restraints used	
NOE restraints*	
<i>All</i>	2641
<i>Protein-ligand</i>	NA
<i>Intermonomer</i>	634 (99)
<i>Intraresidue</i>	980
<i>Sequential (<math> i-j  = 1</math>)</i>	633
<i>Medium (<math>1 &lt;  i-j  \leq 4</math>)</i>	850
<i>Long range (<math> i-j  &gt; 4</math>)</i>	169
Dihedral	
$\phi$ angles	64
$\varphi$ angles	64
Hydrogen bonds	
	90
Structure statistics	
Violations	
<i>Distance (<math>&gt; 0.5 \text{ \AA}</math>)</i>	3
<i>Dihedral angle (<math>&gt; 5^\circ</math>)</i>	1
Energies (cal/mol)	
<i>Overall</i>	-5688 ( $\pm 169$ )
<i>Bond</i>	26 ( $\pm 2$ )
<i>Angle</i>	150 ( $\pm 7$ )
<i>Improper</i>	307 ( $\pm 50$ )
<i>Dihedral</i>	697 ( $\pm 9$ )
<i>Van der Waals</i>	-1293 ( $\pm 12$ )
<i>Electrostatic</i>	-5576 ( $\pm 158$ )
<i>NOE</i>	352 ( $\pm 45$ )
Geometry – average Values	
<i>Bond</i>	$3.45 \times 10^{-3}$ ( $\pm 1.14 \times 10^{-4}$ )
<i>Angle</i>	0.49 ( $\pm 0.018$ )
<i>Improper</i>	1.34 ( $\pm 0.11$ )
<i>Dihedral</i>	40.83 ( $\pm 0.25$ )
<i>Van der Waals</i>	162.96 ( $\pm 12.05$ )
Average pairwise RMSD ( $\text{\AA}$ )**	
<i>Heavy atoms</i>	5.36 ( $\pm 1.13$ )
<i>Heavy atoms – helical region</i>	0.89 ( $\pm 0.09$ )
<i>Backbone</i>	5.75 ( $\pm 1.23$ )
<i>Backbone – helical region</i>	0.32 ( $\pm 0.07$ )
Ramachandran statistics (%)***	
<i>Most favoured regions</i>	90.8 (98.8)
<i>Additional allowed regions</i>	8.8 (1.2)
<i>Generously allowed regions</i>	0.4 (0)
<i>Disallowed regions</i>	0.0 (0)

\*Number in brackets corresponds to the restraints assigned manually

\*\*Helical region corresponds to residues: Glu192-Glu230 and Pro237-Tyr247

\*\*\* Values within brackets correspond to residues Glu192-Glu230 and Pro237-Tyr247 (helical region)

Therefore,  $^{13}\text{C}$  and  $^{15}\text{N}$  resolved NOESY-HSQC were then carefully inspected in order to confirm the existence of these NOEs and consequently the  $\alpha$ -helical region – Appendix (A.1). After proving the structure obeyed this pattern, artificial hydrogen bond restraints were introduced for these regions in order to help the protein folding

during the calculations. The hydrogen bond restraints in addition to some resolved ambiguities during inspection of the  $^{13}\text{C}$  and  $^{15}\text{N}$  resolved NOESY-HSQC finally yielded the expected helical monomer.

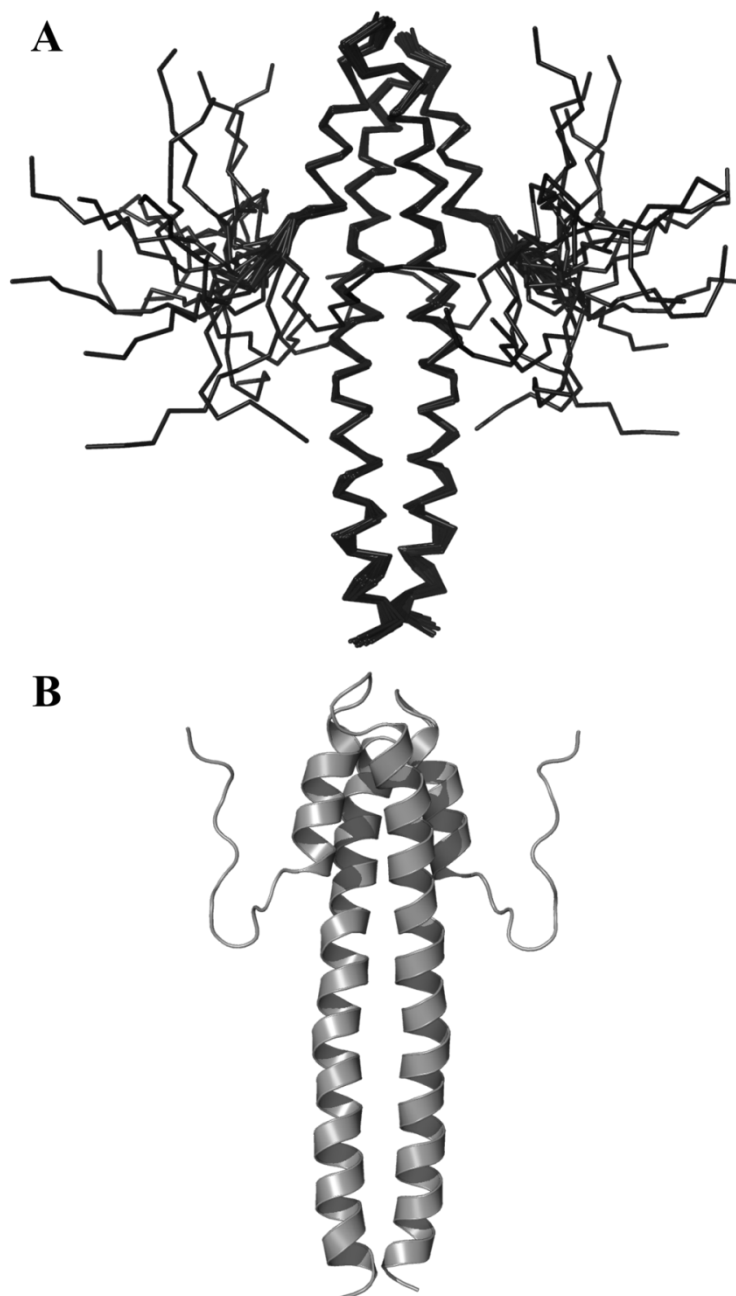


Figure 3.10 – A – Superimposition of the ensemble of 20 lowest energy structures of EB1cΔ8 domain in the free form. B – Cartoon representation of the lowest energy structure for EB1cΔ8.

When compared with heterodimers, homodimers have a larger surface area, fewer hydrogen bonds, higher hydrophobicity and typically  $C_2$  symmetry.<sup>15></sup> The structure determination of symmetric homodimers by NMR is impeded by the fact that is intrinsically impossible to distinguish between inter and intramonomer NOEs. The

only way to resolve this ambiguity is by asymmetric labelling.<sup>152</sup>

As expected, applying C2 symmetry restraints was not enough to obtain the homodimer. The structure obtained was, invariably two parallel monomers, side by side and not a coiled coil. To solve this problem, the inter chain contacts were identified by using inter molecular NOEs from a  $^{13}\text{C}$ ,  $^{15}\text{N}$ -filtered-NOESY-HSQC experiment performed in an isotopically labelled/unlabelled dimer. These NOE restraints were fundamental to determination of the correct structure since they give information on the interface between both monomers. Additionally, CNS protocols were modified in order to optimise these C2 symmetry restraints for a parallel homodimer alternatively to the standard definitions based on an antiparallel homodimeric structure - Figure 3.11.

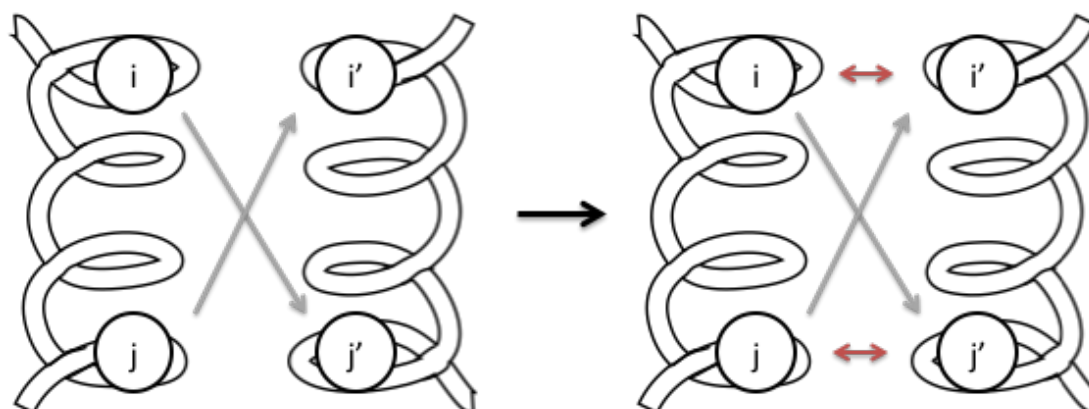


Figure 3.11 – representation of the C2 symmetry restraints used for structure calculation. The 2-fold symmetry axis of the dimer is indicated. The distances from *i* to *j'* and from *j* to *i'* are of equal size. On the left hand side the original restraints (antiparallel) and on the right hand side the modified version (parallel) used in this project.

Because of the leucine zipper structure for the lower region of the helical region, more difficulties arose. As referred before, the methyl region of  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC is highly populated and there are some overlapped peaks, especially the ones belonging to valines and leucines. This is due to the existence of a contiguous heptad repeat in the form *abcdefg*, where *a* is valine and *d* leucine. This repetition of motifs induces similar chemical environments for these side chains (residues Leu196-Leu210) – Figure 3.12.

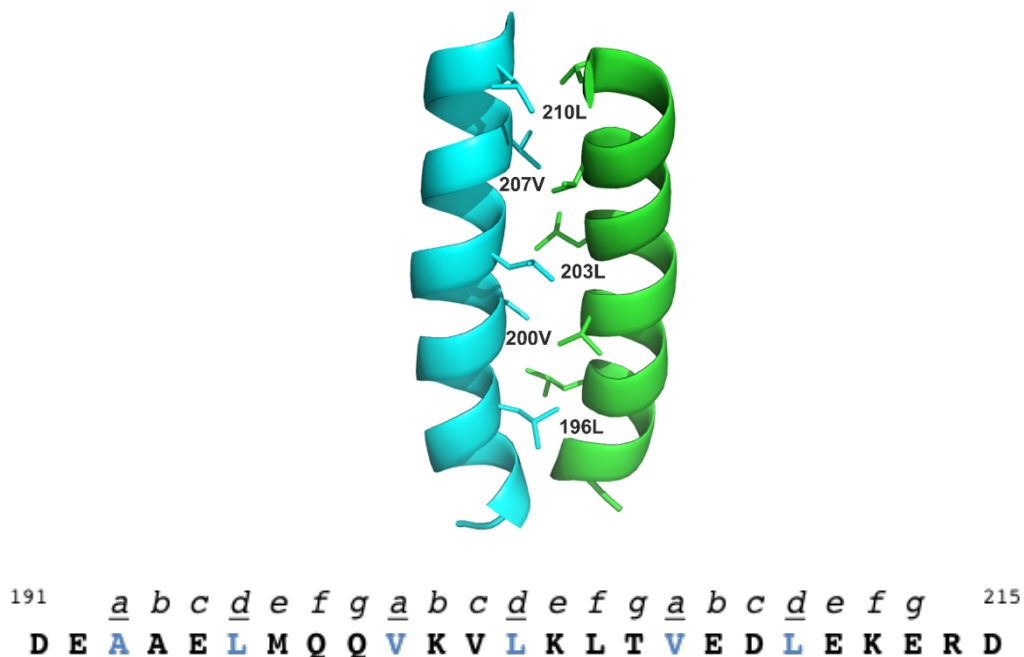


Figure 3.12 - Heptad repeat for the coiled coil structure of EB1c, side chains forming the apolar contacts are shown as sticks.

This helical region structurally characterised as a leucine zipper is immediately followed by a four helical bundle, from Arg214 to Ala248 with a loop (Asn232-Val238) connecting the antiparallel helices for each monomer. The four helix conformation creates additional difficulties for correct assignment and structure determination. Being A and B the two longer helices of EB1c, A' and B' correspond to two identical shorter helices running antiparallel to A/B. Consequently, there are two possible combinations of inter helices contacts: A – A' (intramonomer), A – B' (intermonomer) - Figure 3.13.

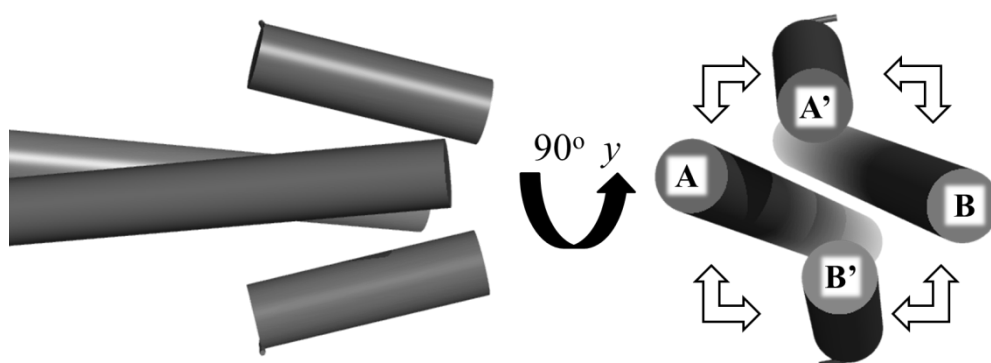


Figure 3.13 – schematic representation of the four helical bundle of the EBH domain of EB1c and possible inter and intramonomer contacts.

This creates additional difficulties as the contacts between A – A' and A – B' are

different, different faces of the helices. If the right interaction surfaces are not correctly assigned, the position of the four helix bundle will be incorrect.

This required detailed analysis of these regions and manual assignment of important contacts. Distance restraints were then derived from this assignment and used in the calculations as absolute NOE restraints, *i.e.* not filtered out during the iterative assignment performed by ARIA,

Outside the fully structured coiled coil region, 248-260, the protein becomes unstructured, being highly dynamic - Figure 3.10A. This trend is supported by the RMSD values – Figure 3.14. The RMSD values for both backbone and side chains increase sharply from Asp250 onwards. Extra support for this flexibility is given by the existence of only intra-residue and sequential (+1/-1) NOE contacts in this region (Figure 3.15) and the intensities of the backbone  $^1\text{H}, ^{15}\text{N}$ -HSQC (Figure 3.5) cross-peaks dramatically increased for the corresponding residues.

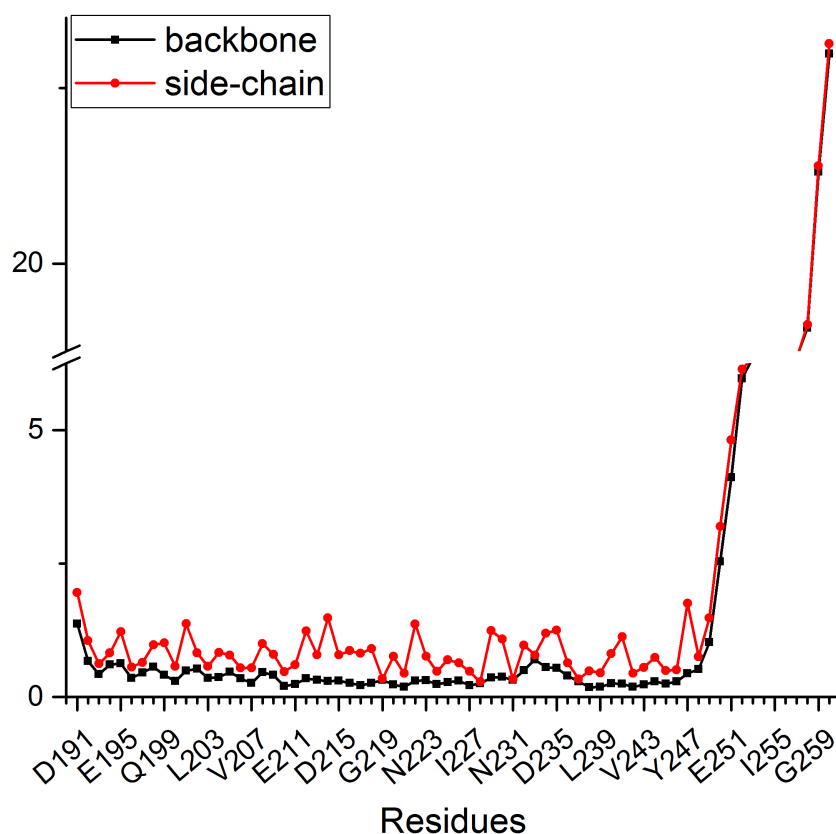


Figure 3.14 – RMSD calculated per residue for the ensemble of 20 structures that form the solution NMR ensemble for EB1cΔ8. In black are shown RMSD values for the backbone and in red for the side chains.

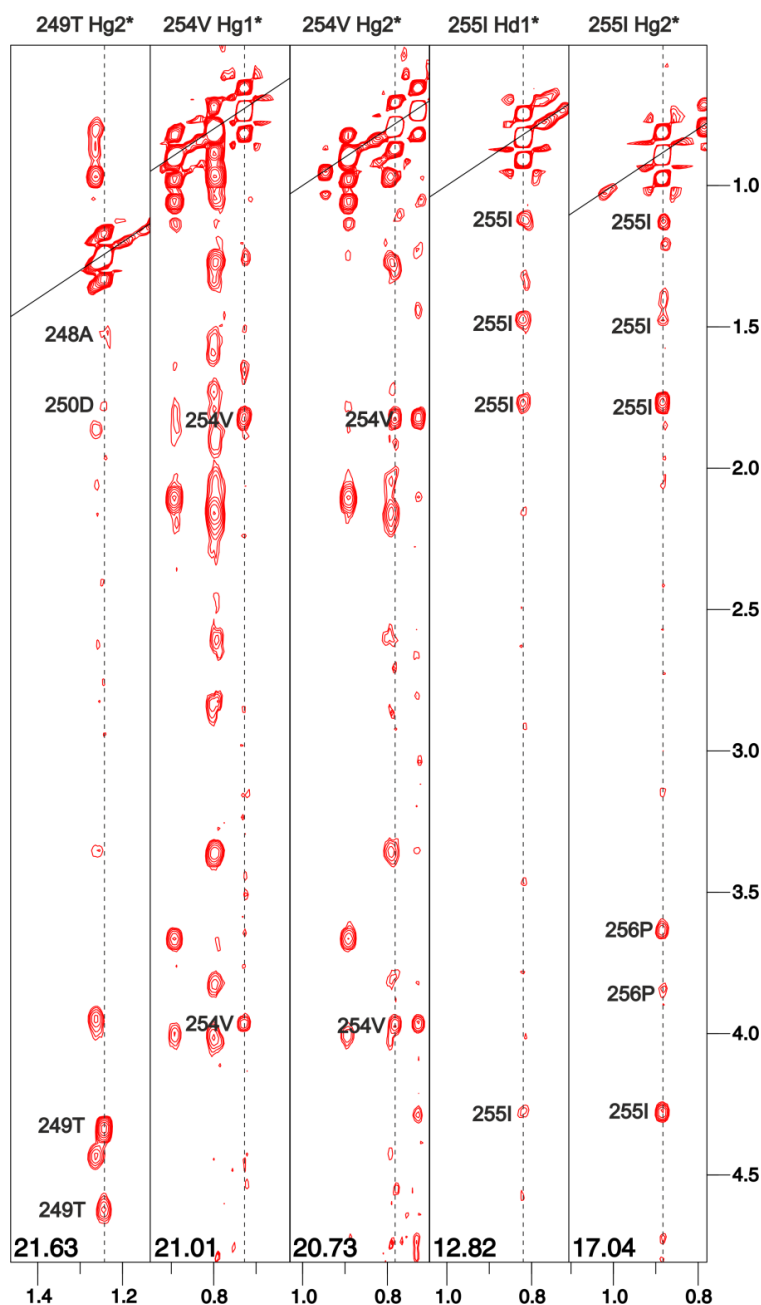


Figure 3.15 – Strips for the  $^{13}\text{C}$ -resolved-NOESY-HSQC for the methyl groups of the following residues of the C-terminus of EB1c $\Delta$ 8 in the free form – Thr249, Val254 and Ile255. The contacts shown refer to the aliphatic region  $\sim 4.5$  ppm to  $\sim 0.4$  ppm and it is possible to observe the absence of inter-residue contacts, except for some sequential residues.

As described in the previous section, 3.3, and shown in Figure 3.9 the aromatic side chain resonances are not very well dispersed, showing some overlap. The main reason for this signal clustering is related with the  $^{216}\text{FYF}^{218}$  motif – Figure 3.16, with Phe216' and Phe218 with similar chemical shifts. The assignment of intermonomer contacts for this region was also of extreme importance since the Phe216' and Tyr217' belong to one monomer, whereas Phe218 belongs to the other. This area is of particular interest

since it forms an aromatic patch at the bottom of the hydrophobic binding site.<sup>43</sup>

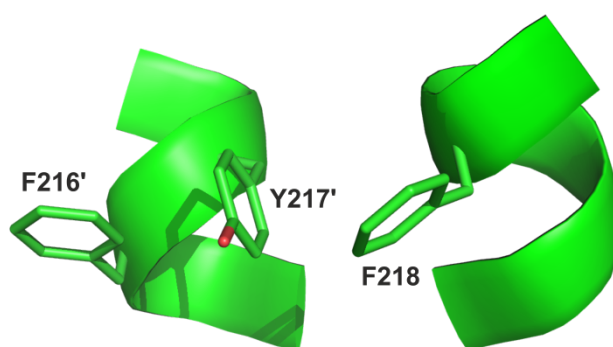


Figure 3.16 – Representation of the aromatic rim of the binding pocket of EB1cΔ8 – the F'Y'F motif.

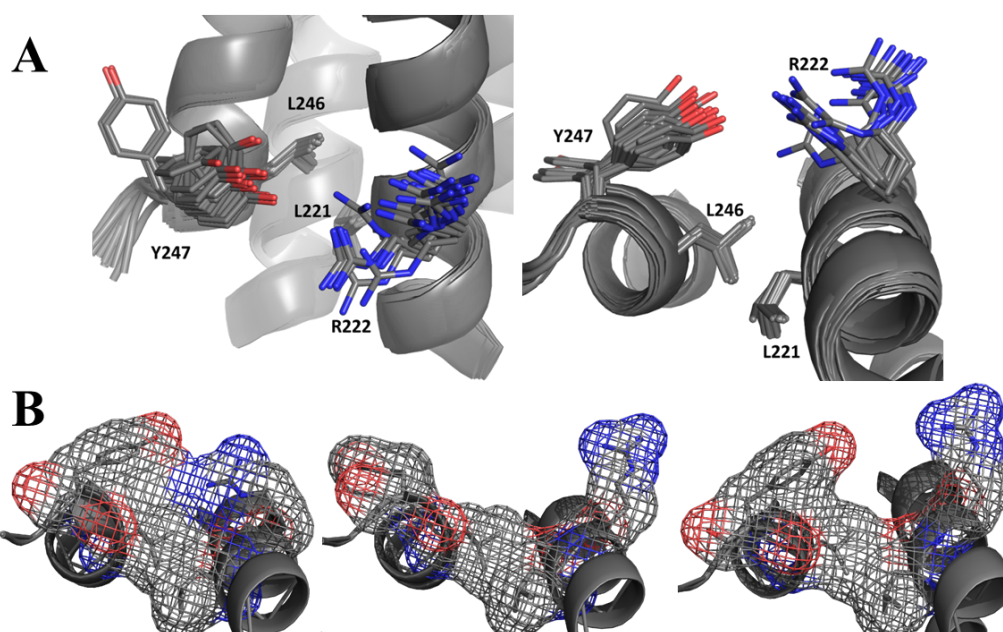


Figure 3.17 - A – conformations for Arg222 and Tyr247 for all the 20 structures calculated for free EB1cΔ8 B - different conformations observed for Arg222 and Tyr247 in solution and how they affect the shape and size of the binding pocket.

It was also observed that two residues part of the hydrophobic pocket, Arg222 and Tyr247, present a dynamic behaviour in solution – Figure 3.17. Side chain RMSD values for the NMR ensemble for these residues are 1.37 Å for Arg222 and 1.76 Å for Tyr247, whereas the average for residues between 193-248 is 0.78 Å. There are NOE contacts between the aromatic ring of Tyr247 and Val243, Asp244, Ile245, Leu246, and Ala248 in the immediate proximity of Tyr247, compatible with all the detected orientations of the side-chain. However, no NOE cross peaks were observed between Tyr247 and Gln240 or Glu225, expected for two of the lowest energy structures of the protein (Figure 3.18). Similarly, there were no NOEs suggesting a preferential conformation of Arg222 side-chain. In the free protein the side-chains of Arg222 and



Tyr247 are dynamic, due to the lack of any specific interactions that would stabilise their orientation.

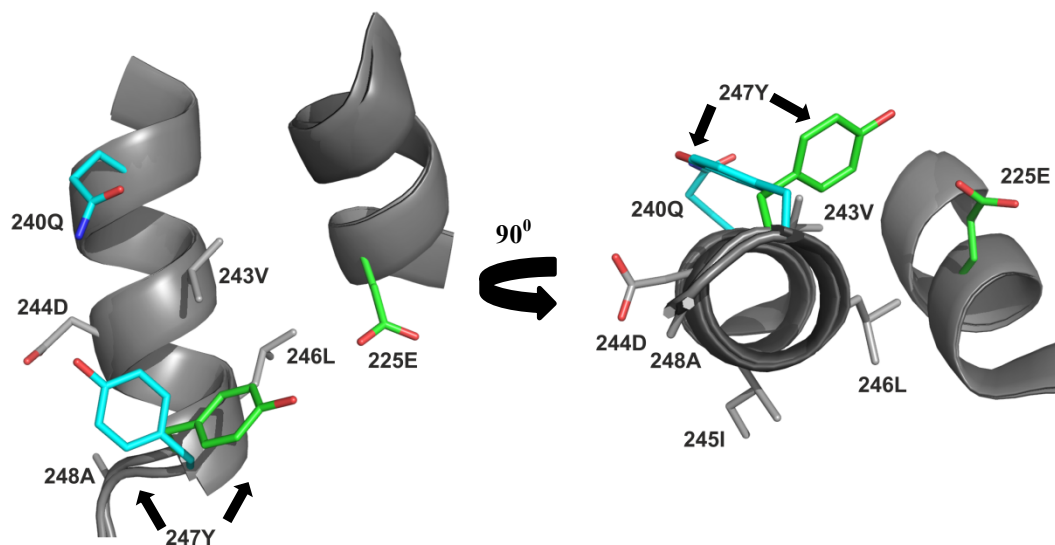


Figure 3.18 – Superimposition of the two lowest energy structures of the ensemble obtained for EB1cΔ8 in the free form. In green it can be seen that Tyr247 is closer to Glu225, whereas in cyan it is closer to Gln240. The remaining residues that show distance restraints to Tyr247 are represented in grey sticks.

Both Arg222 and Tyr247 are part of two of the outer walls of the hydrophobic pocket.<sup>31</sup> Different conformations of these side chains in solution have strong effect on the shape and size of the hydrophobic pocket and therefore on small molecule binding. This was clearly demonstrated in Figure 3.17B, where when both side chains are in an inner position the binding pocket is not formed, as these side chains occlude it. When both point outwards, Figure 3.17B, middle panel, the pocket is partially formed but the outer wall formed by Tyr247 is not defined. When Tyr247 points inwards and Arg222 outwards there is a binding pocket defined where the lateral “walls” are defined by the side chains of these residues.

This matter will be further discussed in section 3.6, but for now the solution NMR structure for EB1cΔ8 in complex with molecule **1a** and differences between free and bound form will be analysed.

### 3.5 Solution NMR structure of the complex EB1cΔ8-1a

In the previous section the solution NMR structure of EB1cΔ8 was elucidated and described. Notable characteristics such as the side chain mobility of residues in the binding site were described. It is important to determine the similarities and differences

between EB1c $\Delta$ 8 in the free form and bound to a small molecule and make conclusions of the main structural changes occurring in EB1c $\Delta$ 8 upon ligand binding. These aspects will be subsequently described.

As described in Chapter 2, the NH resonances for EB1c $\Delta$ 8 upon **1a** titration are affected in a fast exchange manner being possible to follow the chemical shift changes and propagating the assignments. The same was possible for  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC. Most of the resonance assignments were possible to assign using this strategy. The resonances for the small molecule, **1a**, were assigned using  $^{13}\text{C}$ ,  $^{15}\text{N}$ -filtered-NOESY and TOCSY experiments. With all the assignments in place it is possible to calculate distance restraints and perform structure calculation, similarly to what happened with EB1c $\Delta$ 8 in the free form.

Regarding the structure of the complex EB1c $\Delta$ 8-**1a** an additional complication was introduced due to the presence of a ligand. Luckily, it was known from the subtle chemical shift changes in the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC observed upon ligand addition that there were not major structural changes induced by ligand binding, and therefore the structure would not be very different from the free form. Hence, it was decided to use a simplified approach where the final restraints obtained for EB1c $\Delta$ 8 were used as the initial distance restraints input and the intermolecular restraints between the ligand and the protein are introduced to place the ligand relatively to the protein.

So far, for the free protein there were two combinations of intermolecular restraints – chain A to chain B and reciprocally chain B to chain A. Being a symmetric homodimer means also that EB1 will have two symmetrical and equivalent binding sites, both formed by residues from both monomers - A and B. With a ligand bound to each binding site – chains C and D, it is to expect the following combination of intermolecular contacts: *A-B*; *A-C*; *B-C*; *A-D* and *B-D* being a highly complicated system for structure determination using NMR - Figure 3.19. Again, in order to simplify the number of possible intermolecular combinations, only one ligand was introduced in the calculations to simplify the combination of restraints needed. This is a valid strategy as the calculated ligand will be symmetric in terms of binding mode to the second ligand. By doing this, the number of possible combinations for intermolecular restraints is reduced to: *A-C*, *A-B*, *A-C* and *B-C*.

The restraints referring to free EB1c, incompatible with the intermolecular restraints

introduced for the complex were automatically removed by ARIA through filtering of incompatible restraints with the protein-ligand intermolecular restraints. Using this protocol, it was possible to obtain high quality and accurate structures.

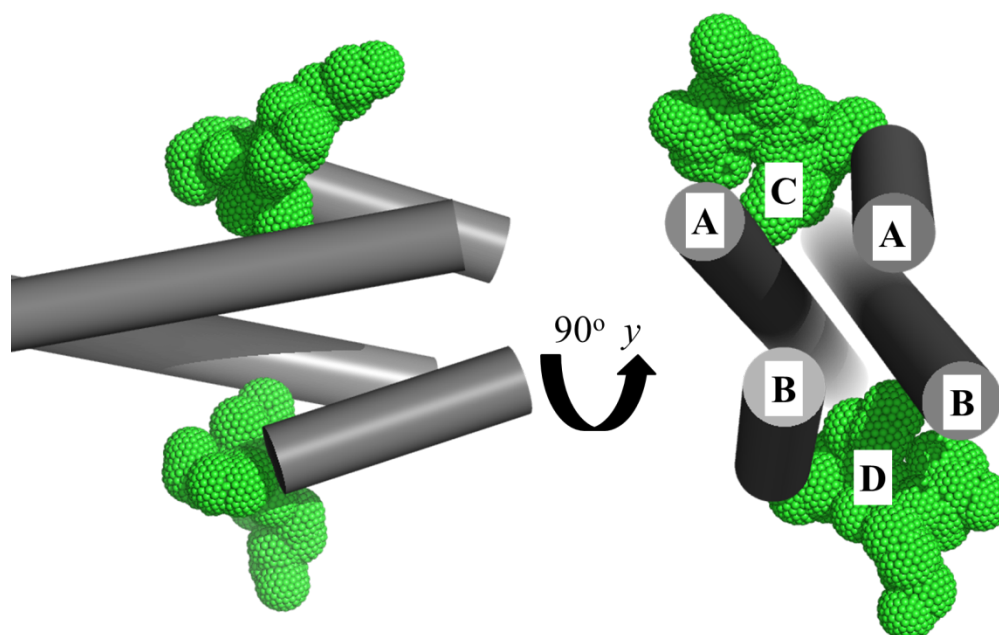


Figure 3.19 – Schematic representation of a ligand bound to EB1cΔ8 binding site. The binding site is formed by both A and B chains and two ligands can bind simultaneously, C and D.

In Table 3.2, the parameters used for the calculation of both structures (free and complex) are shown. In ARIA, the user can choose between Cartesian and Torsion Angle dynamics. Both of these simulating annealing (SA) strategies have been optimised for ambiguous NOE-derived restraints. Torsion angle dynamics generally produces an increased convergence and leads to better local geometries.<sup>15></sup> The parameter “temperature” in this context has no physical meaning, but is simply a measure of the probability of the macromolecule to cross an energy barrier (*i.e.*, its kinetic energy). As mentioned before, to speed up the calculations it was decided to use the lower energy structure obtained for free EB1cΔ8 as the starting structure for the calculation instead of the extended polypeptide chain based only on the sequence of amino acids which is the default option. Therefore, some modifications in the SA parameters used for the calculation were necessary. The temperature used for the complex structure calculation was lower (2000 K) than the one used for calculation of free EB1 (10000 K). Because the structure of free EB1 was used as the initial structure, it was not desirable to use high temperatures that would make the simulations highly dynamic with possible loss of the introduced structure. Instead, the simulations should

have enough freedom only for the protein to adopt the conformation defined by the intermolecular restraints. The number of steps is higher for the structure of the complex since, in general, the larger the number of cooling steps, the higher the percentage of accurate calculated structures, and because the calculated structure was already very close to the final form it was important to make sure the selected parameters would give an accurate structure and satisfy all the intermolecular NOE restraints between the protein and the ligand.

Table 3.2 – Simulating annealing parameters used for structure determination of free EB1cΔ8 and the complex EB1cΔ8/1a.

	Free	Complex
Type	Torsion angle	Torsion angle
Random seed	89764443	89764443
TAD high temperature	10000.0	2000.0
TAD time-step factor	9.0	9.0
Cartesian High temperature	2000.0	2000.0
Cartesian 1 <sup>st</sup> iteration	0	0
Time-step	0.00015	0.00075
Cool1 final temperature	1000.0	1000.0
Cool2 final temperature	50.0	50.0
High-temp steps	20000	40000
Refine steps	8000	16000
Cool1 steps	10000	40000
Cool2 steps	8000	32000

Seventy-five intermolecular NOEs derived from <sup>15</sup>N, <sup>13</sup>C filtered NOESY experiments were used for calculating structure of the complex. Statistics of the structure determination are presented in Table 3.3. This approach yielded a structure of a complex, within a reasonable time and seems a robust and quick way to obtain solution structures of weakly bound complexes. The overall structure is shown in Figure 3.20.

Table 3.3 - NMR distance and dihedral statistics and structure statistics determined for the 20 best structures of the ensemble of the complex EB1cΔ8-1a.

EB1c-1a	
Total restraints used	
NOE restraints*	
<i>All</i>	2766
<i>Protein-ligand</i>	75
<i>Intermonomer</i>	648 (99)
<i>Intraresidue</i>	1044
<i>Sequential (<math> i-j  = 1</math>)</i>	641
<i>Medium (<math>1 &lt;  i-j  \leq 4</math>)</i>	866
<i>Long range (<math> i-j  &gt; 4</math>)</i>	169
Dihedral	
$\phi$ angles	64
$\psi$ angles	64
Hydrogen bonds	
	90
Structure statistics	
Violations	
<i>Distance (<math>&gt; 0.5 \text{ \AA}</math>)</i>	16
<i>Dihedral angle (<math>&gt; 5^\circ</math>)</i>	0
Energies (cal/mol)	
<i>Overall</i>	-5716 ( $\pm 163$ )
<i>Bond</i>	38 ( $\pm 1.5$ )
<i>Angle</i>	167 ( $\pm 5.5$ )
<i>Improper</i>	298 ( $\pm 32$ )
<i>Dihedral</i>	713 ( $\pm 7$ )
<i>Van der Waals</i>	-1298 ( $\pm 10$ )
<i>Electrostatic</i>	-5635 ( $\pm 17$ )
<i>NOE</i>	1280 ( $\pm 50$ )
Geometry – average Values	
<i>Bond</i>	$4.65 \times 10^{-3}$ ( $\pm 1.18 \times 10^{-4}$ )
<i>Angle</i>	0.67 ( $\pm 0.013$ )
<i>Improper</i>	1.31 ( $\pm 0.071$ )
<i>Dihedral</i>	40.94 ( $\pm 0.25$ )
<i>Van der Waals</i>	346.16 ( $\pm 23.53$ )
Average pairwise RMSD ( $\text{\AA}$ )**	
<i>Heavy atoms</i>	4.54 ( $\pm 9.28 \times 10^{-1}$ )
<i>Heavy atoms – helical region</i>	$6.12 \times 10^{-1}$ ( $\pm 7.73 \times 10^{-2}$ )
<i>Backbone</i>	4.92 ( $\pm 1.12$ )
<i>Backbone – helical region</i>	$1.57 \times 10^{-1}$ ( $\pm 3.61 \times 10^{-2}$ )
Ramachandran statistics (%)***	
<i>Most favoured regions</i>	90.6 (98.1)
<i>Additional allowed regions</i>	9.3 (1.9)
<i>Generously allowed regions</i>	0.1 (0)
<i>Disallowed regions</i>	0.0 (0)

\*Number in brackets corresponds to the restraints assigned manually

\*\*Helical region corresponds to residues: Glu192-Glu230 and Pro237-Tyr247

\*\*\* Values within brackets correspond to residues Glu192-Glu230 and Pro237-Tyr247 (helical region)

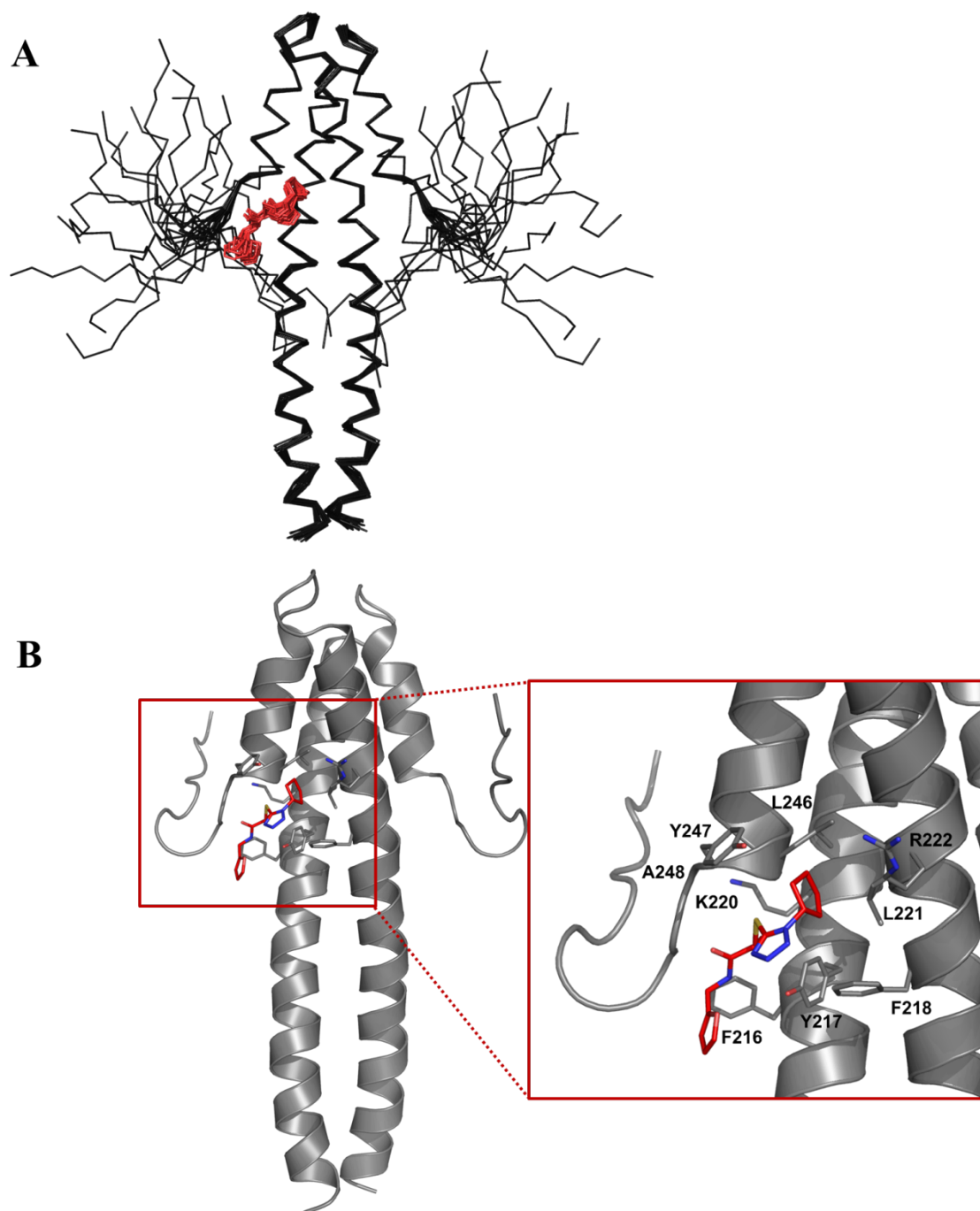


Figure 3.20 - A – Superimposition of the ensemble of 20 lowest energy structures of EB1c $\Delta$ 8 domain in complex with **1a**. B – cartoon representation of the lowest energy structure for EB1c $\Delta$ 8 in complex with **1a**. **1a** is shown in sticks, where red represents carbon, blue nitrogen and yellow sulphur.

Analysis of the structure reveals two important features, the first is compound **1a** binds to the same region reported to be targeted by the natural SxIP motif and second the fact that this structure is very similar to the free form. This last fact is also supported by the side chain RMSD values obtained for both ensembles and shown in Figure 3.21. The C-terminus remains largely flexible as shown by Figure 3.20A. Similarly, to what

was observed for free EB1c $\Delta$ 8, RMSD values *per* residue across the 20 structures that compose the solution NMR ensemble for this complex show a sharp increase in these values is observed from Asp250 onwards – Figure 3.21. Moreover, the only residue outside of the structured region that has an intermolecular NOE cross-peak is Thr249, which occupies a region very close to the coiled coil.

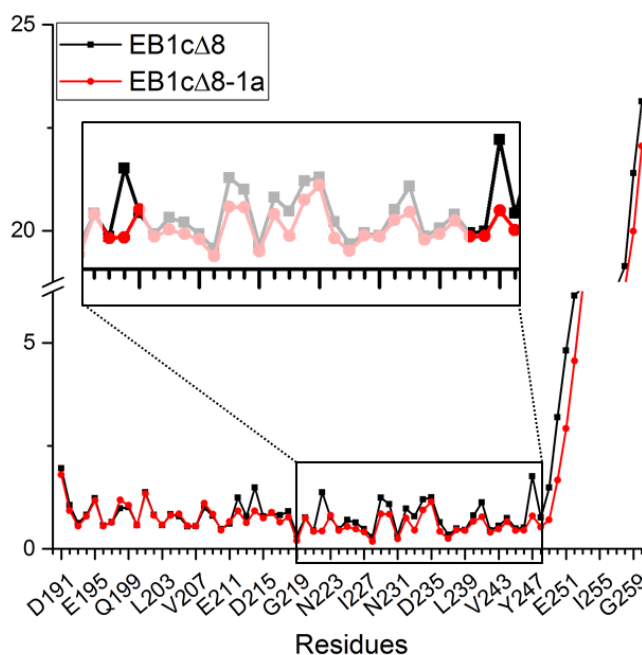


Figure 3.21 – RMSD values for EB1c $\Delta$ 8 side chain for free (black) and bound to **1a** (red) structures.

In terms of intermolecular distance restraints between **1a** and EB1c $\Delta$ 8 in the final iteration they position the cyclopentyl ring close to Tyr217, Phe218, Leu221, Arg222, Leu246 and Tyr247 – all consistent with the hydrophobic binding site targeted by the IP motif of the SxIP proteins. The methyl group shows NOE cross-peaks to Tyr217, Phe218 and Tyr247, and additional contacts are made with Phe216 and Thr249, the latter outside the coiled coil region. Finally, the oxazole moiety does not show any intermolecular NOEs. CSP and molecular docking data – Chapter 2 - are, therefore, in agreement with the three-dimensional structure of the complex.

Regarding the binding site dynamics, Arg222 and Tyr247 are now in a more fixed conformation when compared with the free form – Figure 3.22 (compare with Figure 3.17). Side chain RMSD values are now 0.43 Å for Arg222 and 0.80 Å for Tyr247, whereas previously were 1.37 Å and 1.76 Å, respectively. Figure 3.21 shows a comparison of the RMSD values for amino acid side chains between EB1c $\Delta$ 8 and EB1c $\Delta$ 8-**1a** complex and the larger RMSD values for both Arg222 and Tyr247 can be

observed.

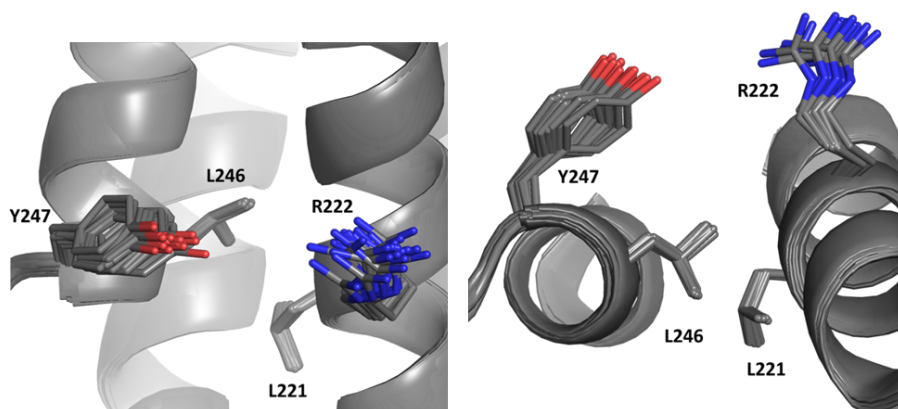


Figure 3.22 – conformations for Arg222 and Tyr247 for all the 20 structures calculated for free EB1c $\Delta$ 8 bound to compound **1a**.

The existence of intermolecular NOEs observed between these residues and the cyclopentyl ring of **1a**, Figure 3.23, confirms that **1a** has a stabilising effect on these side-chains. As consequence, the binding pocket remains in the same conformation for the ensemble of structures obtained for this complex.

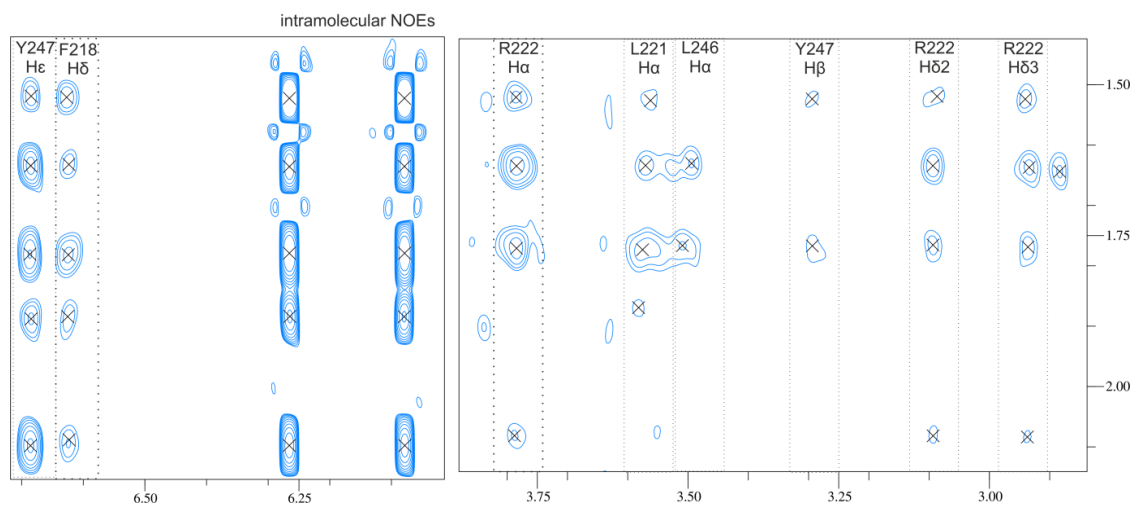


Figure 3.23 - Selected regions of a  $^{13}\text{C}$ ,  $^{15}\text{N}$  filtered-2D-NOESY showing intra and intermolecular NOEs observed for EB1c-**1a** complex, acquired at 25°C. The intramolecular NOEs correspond to the aromatic protons of the oxazole moiety NOEs to the cyclopentyl ring. The intermolecular NOEs show the NOE contacts between the cyclopentyl ring and Phe218, Leu221, Arg222, Leu246 and Tyr247.

The stabilisation of EB1's binding pocket upon ligand binding has not been reported before. However, analysis of published X-ray crystal structures obtained for free and bound form of EB1c indicate that Arg222 and Tyr247 adopt two different conformations in the free (PDB codes 1YIG and 1WU9<sup>27,43</sup>) and bound state, (PDB code 3GJO<sup>31</sup>) – Figure 3.24A.



These conformations clearly affect the shape of the binding site – Figure 3.24B. In the absence of a ligand the binding pocket is dynamic, changing from the fully open state to the fully closed state (Figure 3.17B) with the open state characterised by Tyr247 side-chain pointing outwards and Arg222 side-chain flattened against the surface of the helix leading to the absence of one of the outer walls of the binding pocket. This structure is stabilised in the crystallised free EB1c (Figure 3.24B). In the closed form, side-chains of Arg222 and Tyr247 point towards each other, completely blocking the binding site. This closed form is present in the NMR ensemble for the free protein (Figure 3.17B, top panel). Finally, the binding pocket is fully formed in the crystal structure of the complex (Figure 3.24B), complex with **1a** and some of the structures of the free protein NMR ensemble (Figure 3.17B, bottom panel).

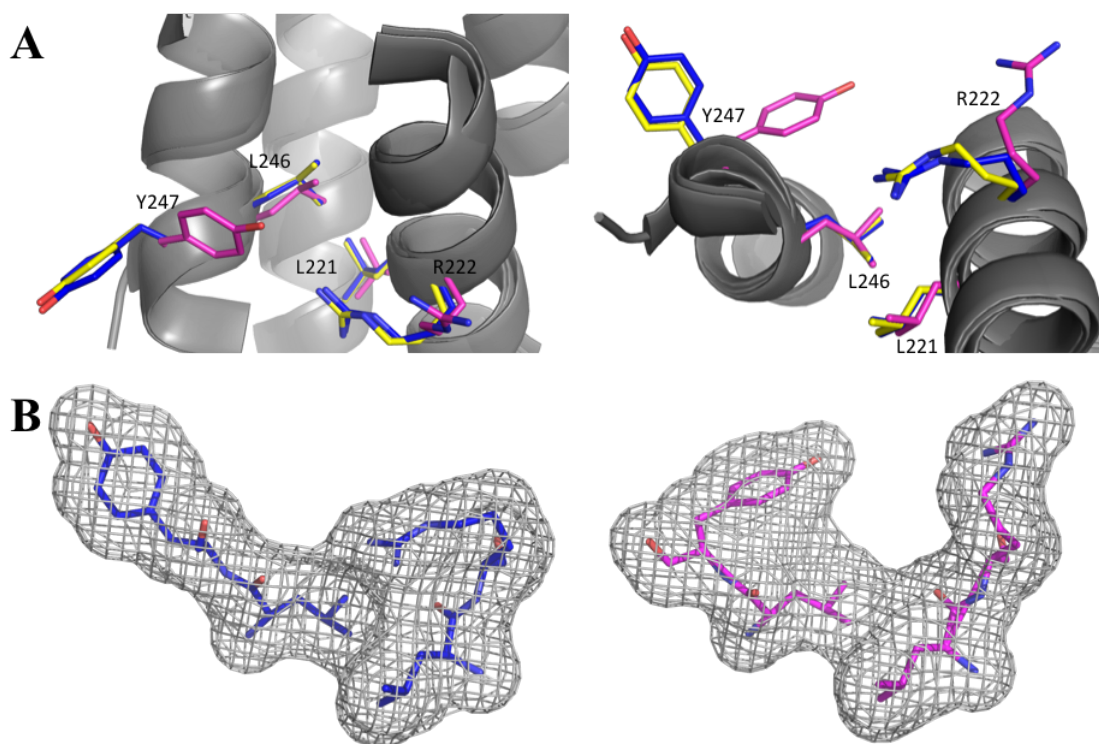


Figure 3.24 - A - Two conformations for residues Arg222 and Tyr247 in free (yellow and navy, PDB codes 1YIG and 1WU9 respectively) and bound state (magenta, PDB code 3GJO). Leu221 and Leu246 remain in a stable conformation as they are part of the coiled coil hydrophobic interface and are shown as reference points. On the right hand side the same representation with a 270° rotation on the *x* axis. B – Representation of the EB1 pocket shape based on the conformation of residues Arg222 and Tyr247, for unbound EB1c (PDB code 1YIG), and bound EB1c (PDB code 3GJO).

### 3.6 Analysis of the implications of the side chain dynamics for Arg222 and Tyr247 and its influence on small molecule binding

To further explore the potential effects of the binding pocket dynamics a simple experiment was designed where small molecule binding to EB1c $\Delta$ 8 was re-evaluated using a variety of EB1c structures with different binding site conformations. It has been widely reported that the cross-docking of a ligand to an ensemble of protein structures increases the chances to find near-native solutions.<sup>89</sup> Cross-docking consists of docking a ligand to an ensemble of different protein conformations.<sup>15</sup> This approach is essential when targeting dynamic regions.<sup>155</sup>

#### 3.6.1 Cross-docking approach

The four experimentally tested molecules were re-docked using the ensemble of 20 NMR structures for the free form and three crystal structures – 1YIG<sup>27</sup>, 1WU9<sup>43</sup> and 3GJO<sup>31</sup>. The first two are a three-dimensional model of EB1c in the absence of ligands whereas the latter represents EB1c bound to a SxIP motif containing peptide.

Average docking score for the NMR ensemble for compound **1d** was highest, 64, followed by **1a**, 53, **1c**, 48 and **1b** with a score of 46. This ranking is in accordance with the *in vitro* screening, where **1d** is the best binding compound, followed by **1a**; compounds **1c** and **1b** have much weaker interactions. In contrast, when docked to the crystal structure of complex with the peptide (3GJO), compound **1b** had a score of 59 that was higher than for compound **1a** (55), in clear disagreement with the experimental measurements. This suggests that docking to the NMR ensemble that accounts for the dynamics of the binding site, providing a more reliable prediction of the binding propensity than the docking to the fully formed binding site of the complex.

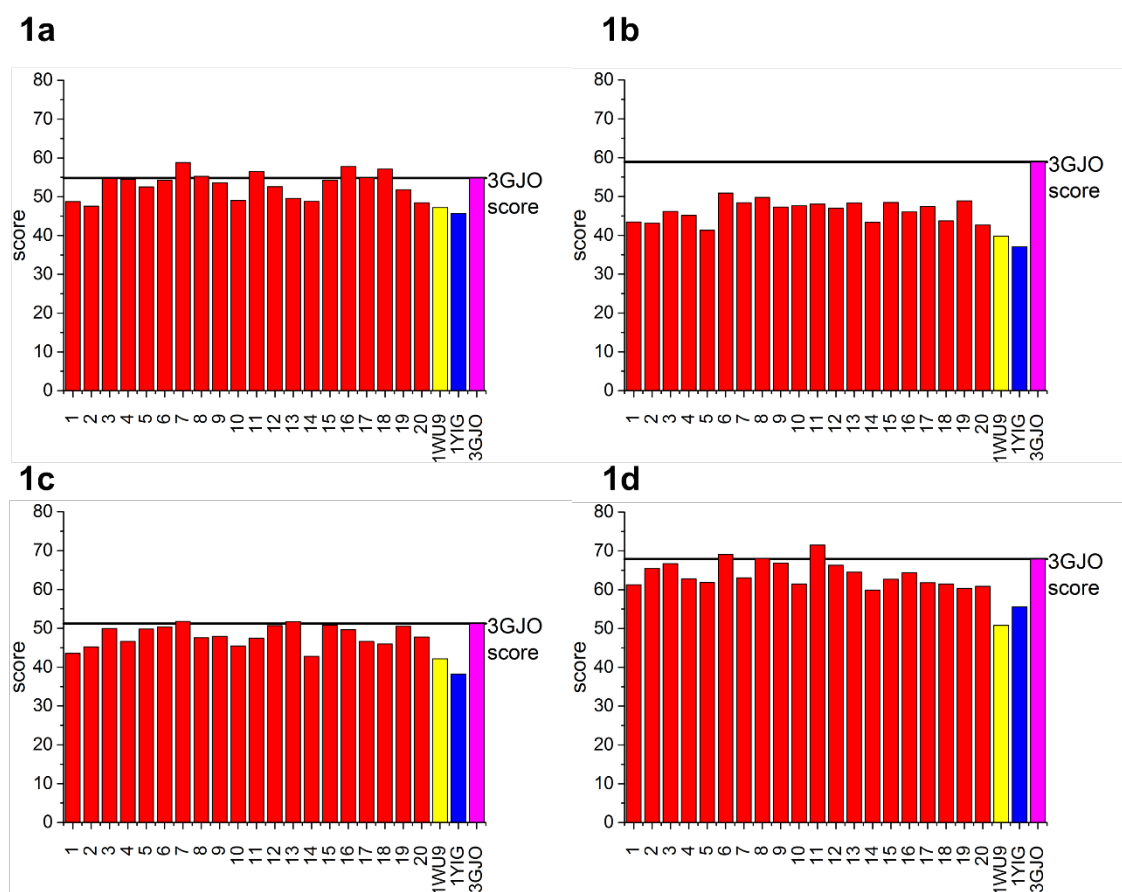


Figure 3.25 - Predicted docking scores for each compound, **1a**, **1b**, **1c** and **1d**, for the ensemble of the solution NMR structures of free EB1 (red), crystal structures of free EB1, 1WU9 (yellow) and 1YIG (navy), and crystal structure of EB1 in the bound state with SxIP protein, 3GJO (magenta). The score obtained for the reference structure where the initial docking calculations were performed – 3GJO, is defined by a black horizontal line to facilitate the comparison.

The modelling indicates that compounds **1a** and **1d** can interact with a wide range of states, where binding site is partially formed, while **1b** and **1c** only bind to the fully formed binding pocket. In agreement with this, while all compounds had low scores when docked to the open binding pocket of the crystal structures of the free EB1c, the scores of the compounds **1a** and **1d** were significantly higher than the scores of the other two compounds (Figure 3.25). The docking results provide an explanation of why compound **1b**, that apparently fits the binding pocket well, shows negligible interaction with EB1c. This compound can only interact with an extremely small population of EB1c where the binding site is fully formed spontaneously, while compounds **1a** and **1d** interact reasonably well with the majority of the configurations of the binding pockets, potentially inducing further binding pocket changes after an initial docking. Closer agreement with the experimental results suggests that the use of solution NMR structures and cross-docking can be a powerful tool in drug design

for dynamic regions.

### 3.6.2 Principal component analysis

Principal component analysis (PCA) is a method to reduce the dimensionality of a data set to facilitate a quicker analysis of various variables at the same time, retaining the information contained in these variables. This allows for the conversion of possibly correlated variables into a set of linearly uncorrelated variables (principal components) that can give an insight into the relationships between variables.<sup>155</sup> The first principal component explains the maximum variance in the data set, the second component describes the maximum of the remaining variance, and so on. Therefore, the majority of the information in most of data sets will be contained in the first few principal components.<sup>156</sup> To relate the variations in the binding site geometry to the ligand binding properties, a set of geometric parameters that characterise binding site variation were calculated and used in PCA to isolate and compare the main configurations of the binding pocket.

Since the main changes in the IP binding pocket are attributed to the Tyr247 and Arg222 side-chain dynamics, the distances between  $\zeta$  groups of these residues, as well as between  $\zeta$  of Arg222 and  $\zeta$  of Tyr247 to the invariable  $\gamma$  groups of Leu221 and Leu246 respectively were measured. Additionally, CASTp<sup>157</sup> was used to identify and characterise the binding pockets, in terms of size (area and volume), number of openings, sum arc length and corner points.

Table 3.4 - Parameters used for PCA analysis for the hydrophobic pocket shape and size analysis. NMR ensemble for free EB1cΔ8 is represented by *a* followed by a number where 1 is the lowest energy structure and 20 the highest energy structure. For the complex, EB1cΔ8-1a, the same approach was used but the letter *b* was used instead.

structure	Distances in Å			Number of mouth openings	CASTp calculated parameters				sum arc length	corner points
	Y247-L246	Y247-R222	R222-L221		Area (Å <sup>2</sup> )		Volume (Å <sup>3</sup> )			
					Solvent accessible	Molecular surface	Solvent accessible	Molecular surface		
a1	6.2	9.5	10.8	0	1.01	39.76	0.047	22.54	5.95	10
a2	5.7	7.7	8.1	0	8.511	83.92	0.834	54.98	23.73	24
a3	5.7	9	10.8	1	55.546	117.99	28.662	146.9	61.14	35
a4	5.7	6.6	8.7	1	52.497	120.81	25.188	143.57	62.64	34
a5	8.2	12.9	8.2	1	17.126	98.98	2.151	77.16	36.74	29
a6	5.5	6.8	7.8	1	52.621	133.46	23.978	150.3	65.33	44
a7	5.8	9.3	10.8	2	235.65	364.76	222.704	627.24	196.12	96
a8	5.3	8.4	10.9	1	48.999	140.17	21.852	148.61	70.47	44
a9	6.2	10.1	11	1	43.184	105.82	18.735	116.85	46.96	31
a10	5.7	7.7	9.3	1	64.44	121	41.899	170.09	68.49	38
a11	8.2	15.5	11	2	62.05	140.35	29.281	164.23	71.83	41
a12	5.9	9.8	10.7	1	45.449	102.4	25.231	126.74	51.68	33
a13	6.9	10.1	10.9	1	61.746	141.29	33.273	171.06	72.64	46
a14	5.5	8.5	10.1	1	34.944	95.88	12.706	99.36	41.04	27
a15	5.4	8.5	10.1	1	48.167	129.3	18.606	138.81	57.76	42
a16	8.1	15.9	10.3	1	59.374	123.11	33.243	159.51	69.51	39
a17	6.1	6.9	8.1	1	48.32	107.57	22.138	128.66	56.61	32
a18	5.9	6.1	7	1	60.851	128.31	29.974	154.73	59.46	39
a19	6.4	6.6	8.1	1	47.268	122.42	22.083	136.88	56.2	40
a20	5.7	8.5	10.6	1	50.097	177.66	20.751	165.42	71.37	45

Structural characterisation of EB1c using solution NMR

structure	Y247-L246	Y247-R222	R222-L221	Number of mouth openings	Area (Å <sup>2</sup> )		Volume (Å <sup>3</sup> )		sum arc length	corner points
					Solvent accessible	Molecular surface	Solvent accessible	Molecular surface		
b1	7.1	8.7	9.9	2	113.463	260.95	63.765	310.72	131.55	81
b2	6.7	7.9	9.8	2	67.6	152.44	43.335	192.84	75.1	47
b3	7	9.4	9.9	2	73.991	202.17	40.768	228.6	100.62	66
b4	6.8	7.6	9.6	2	85.914	235.13	45.588	264.4	116.44	73
b5	6.6	8	9.7	2	88.005	233.83	44.537	265.02	119.3	70
b6	6.6	8.4	9.8	2	95.974	240.12	42.286	270.55	122.06	80
b7	CASTp could not find the right pocket for this structure. It was, therefore, removed from the statistical analysis.									
b8	6.9	8.8	9.8	2	91.69	239.29	50.976	255.84	109.45	68
b9	6.3	7.8	9.8	2	74.923	209.82	48.153	241.05	108.69	73
b10	7.2	7.6	9.4	2	84.547	226.73	52.558	260.92	112.61	74
b11	7.1	8.8	9.8	2	86.247	230.54	50.091	260.88	108.07	74
b12	7	8.7	9.7	2	88.183	197.39	52.529	249.95	108.17	59
b13	7	8.2	9.7	2	90.155	262.6	46.776	275.84	114.23	71
b14	6.6	8.2	9.8	2	84.244	232.91	45.891	252.99	104.37	78
b15	6.6	8.2	9.8	2	78.433	215.94	43.006	234.72	106.5	63
b16	6.4	8	9.8	2	102.369	238.6	56.804	289.47	135.06	79
b17	6.5	8	9.9	2	79.677	213.7	48.089	247.09	107.11	67
b18	7	9.9	9.9	2	72.998	206.1	40.851	230.86	101.43	66
b19	7	9.1	9.9	2	104.845	257.13	53.173	106.03	52.39	33
b20	6.9	8.9	9.9	2	91.618	232.73	58.431	279.48	112.98	71
1wu9	10.9	11.4	5.5	1	12.183	34.72	4.811	34.67	14.53	12
lyig <sup>a</sup>	11.1	11.5	5.5	0	0	0	0	0	0	0
3gio	5.7	9.2	9.9	1	161.483	248.75	123.198	409.01	138.15	70

<sup>a</sup>CASTp could not find any pocket for the second for the closed conformation for this structure.

These distances and calculated parameters are described in Table 3.4 and were then used to characterise the overall binding site geometry through PCA analysis.

The first two components of PCA accounted for 86% of the total variation, demonstrating that they isolate the main binding site changes and can be reliably used for comparison. Notably, the distances of the Tyr247 and Arg222 side-chains have a larger contribution for PC2, while the overall binding site parameters calculated via CASTp for each pocket have a larger contribution for PC1 - Table 3.5. Thus, the location of the structure in the PC1, PC2 plane reflects the relationship between the state of the binding pocket and the orientation of the variable side-chains.

Table 3.5 - PCA analysis results, including standard deviation, variance explained by each principal component and total variance explained. The following rows show the PCA loadings of each parameter to each principal component.

Property	PC1	PC2	PC3
Standard Deviation	2.5422	1.4516	0.94879
Variance Explained	0.64627	0.21071	0.090019
Total Variance Explained	0.64627	0.85698	0.94700
Y247-L246	-0.015618	0.61203	0.43986
Y247-R222	0.032717	0.67285	-0.021162
R222-L221	0.10616	0.32795	-0.87765
Number of mouth openings	0.30574	0.22264	0.17339
Area_SA	0.38862	-0.050398	0.015697
Area_MS	0.38557	-0.055073	0.020567
Volume_SA	0.37779	-0.058885	-0.013655
Volume_MS	0.38932	-0.054128	0.0087726
sum arc length	0.39108	-0.027500	0.028714
corner points	0.38342	-0.053330	0.063138

PCA analysis shows good separation between crystal structures of the free (navy and yellow) and peptide-bound (magenta) EB1c on both PC1 and PC2 axis, in agreement with the large differences in the binding site configuration. In the crystal structure of the free form the binding site is not formed, corresponding to low volumes of the pocket ( $5 \text{ \AA}^3$  for 1WU9) and low PC1 values. Side-chain of Tyr247 for this structure is pointing away from the binding pocket, with large distances and large values of PC2. In contrast, the binding pocket is fully formed for the crystal structure of the complex with the peptide, corresponding to large pocket volumes ( $123 \text{ \AA}^3$ ) and close distances between the side-chains of Tyr247 and Arg222, thus large PC1 and small PC2 values.

These structures provide references for the two opposite states of the binding pocket.

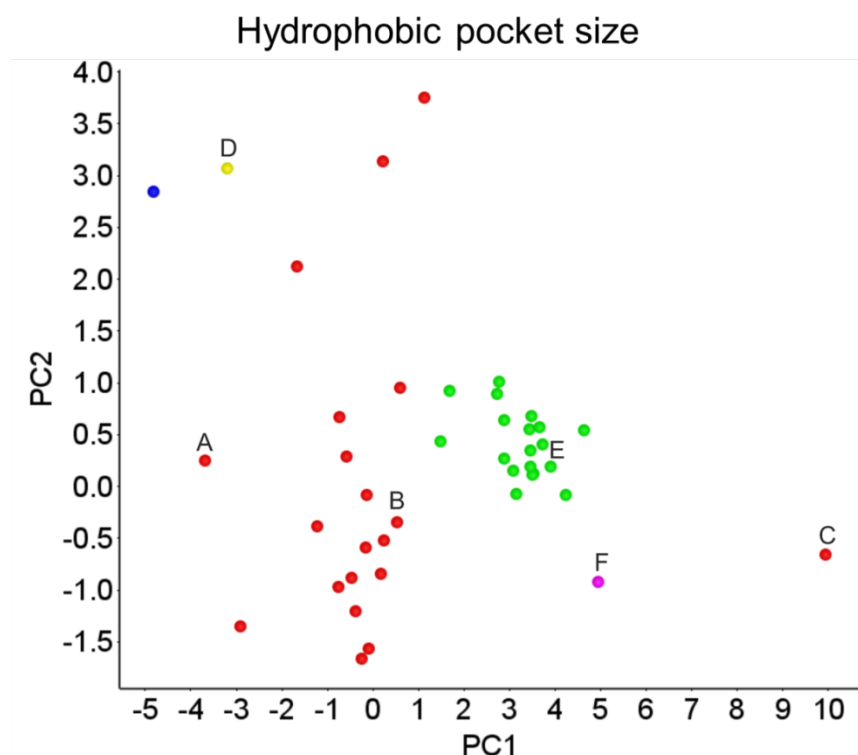


Figure 3.26 - PCA analysis of a set of parameters collected to describe the EB1c binding pocket conformation. The analysis combines solution NMR structures and crystal structures – solution NMR structures for unbound form of EB1c are shown in red, for the complex with **1a** are shown in green, 1WU9 in yellow, 1YIG navy, 3GJO pink. (A total of 86% of the variance can be explained by the two principal components, PC1 and PC2 –Table 3.5).

The structures of the complex with **1a** (green) cluster close to the parameters for the peptide-bound conformation (3GJO, pink). This demonstrates that the binding pocket not only adopts a similar shape and size to the one induced by the natural ligand, but Tyr247 and Arg222 are in similar positions. The average volume of the pocket for the bound to **1a** NMR conformations is  $49 \text{ \AA}^3 (\pm 6 \text{ \AA}^3)$ , reduced compared to the crystal structure of the peptide complex.

The structures of the NMR ensemble of free EB1c in solution demonstrate the largest variation in the binding site characteristics. The majority of structures are located in the vicinity of the complex with **1a**, however a number of structures are spread across the plane. The size of the binding pocket shows a high variability, changing from  $0.04 \text{ \AA}^3$  for the closed state to  $223 \text{ \AA}^3$  corresponding to the transiently folded C-terminus, with the average of  $32 \text{ \AA}^3 (\pm 45 \text{ \AA}^3)$ . This variability is illustrated by the three representatives of the free form of EB1c (Figure 3.27, panels A, B and C), alongside



with three representatives of free and bound EB1c, (Figure 3.27, panels D, E and F).

From the reduced dynamics of Tyr247 and Arg222 side-chains and their optimal orientation in the complex with **1a**, the interactions at the IP binding site are largely utilised by the **1a** scaffold, in agreement with the binding properties observed for the natural SxIP ligand. The identified IP motif mimetic can be used as a starting point for the design of more potent inhibitors to target EB1.

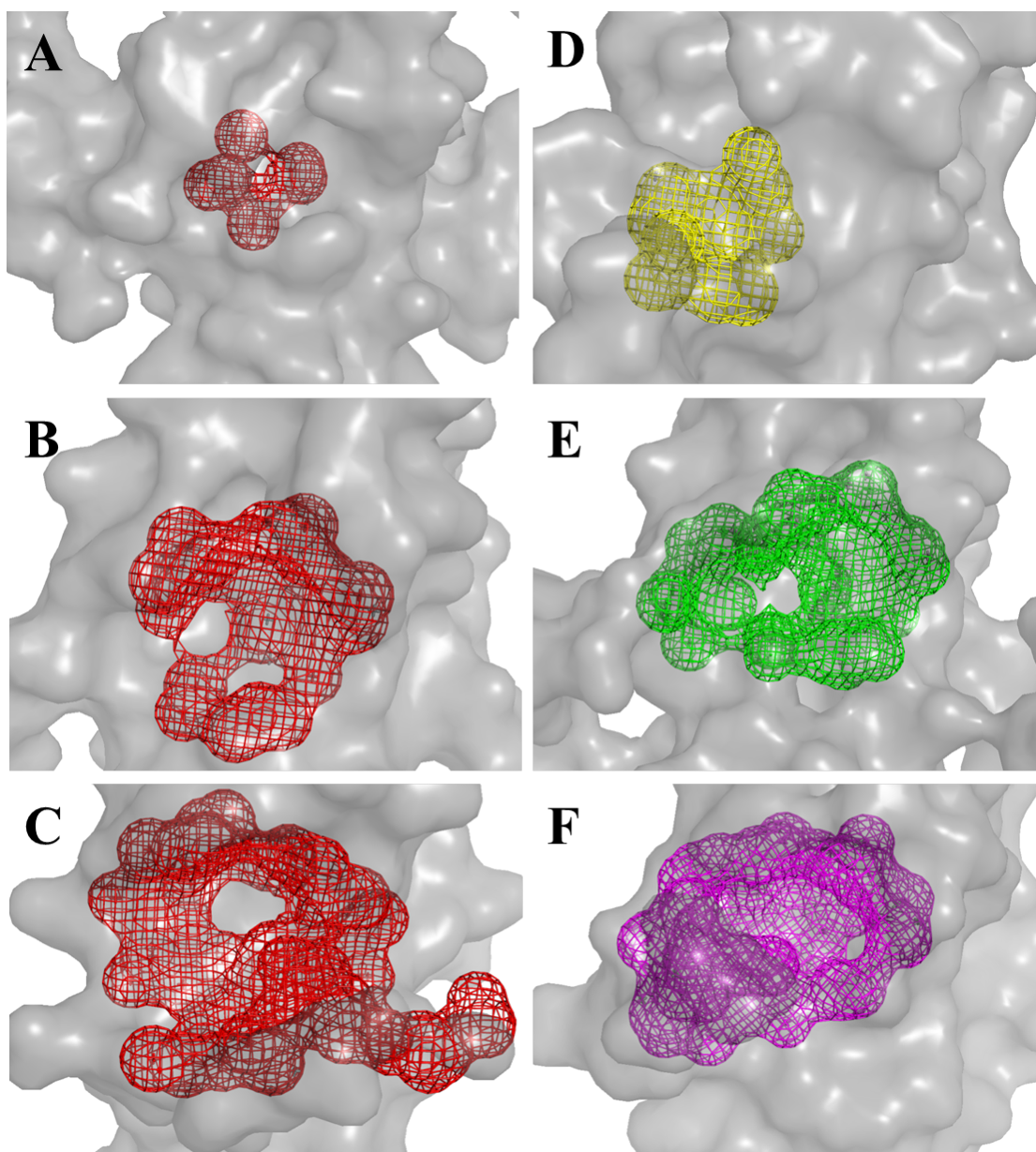


Figure 3.27 - Representation of the binding pockets calculated by CASTp<sup>15</sup> webserver – A, B and C, structures 1, 10 and 7 from the NMR ensemble for free EB1cΔ8. D, E and F 1WU9, structure 6 from NMR ensemble of the bound form and F 3GJO.

However, the volume of the binding pocket of the crystal structure with the SxIP containing peptide indicates that stabilisation of the Arg222/Tyr247 is not enough to maximise the size of the hydrophobic pocket. For the complex with the native SxIP containing peptide (structure 3GJO<sup>31</sup>) the C-terminal flexible region seems to fold on the ligand protecting it from the solvent and expanding the hydrophobic pocket. Thus, the next generation of the inhibitors should be aimed at extending the binding pocket by engaging the dynamic C-terminus and achieve a tighter interaction. That will be the focus of the next Chapter.

## Chapter 4. DESIGN AND IDENTIFICATION OF INHIBITORS BEYOND SxIP MOTIF

Results obtained from the first generation of inhibitors suggested that it was possible to target EB1c $\Delta$ 8 with small molecules based on IP motif of SxIP proteins. Subsequently, in Chapter 3, it was demonstrated that this scaffold stabilises the hydrophobic pocket of EB1 that is dynamic in solution.

However, this did not appear to be sufficient to achieve a strong interaction with EB1c. Ideally, one would like to keep part of the scaffold shared between compounds **1a** and **1d**, the 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide – Figure 4.1A, that is thought to replace IP motif, and further extend the molecule in order to gain further interactions and achieve tighter binding. Therefore, the aim of this chapter will be the design and identification of molecules that target the SxIP binding region and simultaneously extend the contact network outside this region.

### 4.1 Rational design of a hybrid molecule combining the SxIP mimetic scaffold and a tri-peptide

In order to identify which interactions beyond the SxIP motif are important, the crystal structure published by Honnappa et al.<sup>31</sup> (PDB code 3GJO) was again examined, Figure 4.2. These authors reported that the heptapeptide <sup>5476</sup>PSKIPTP<sup>5482</sup> is invariably bound in very similar conformations, and residues outside this core segment are less well defined and do not participate in specific intermolecular interactions. Moreover, Thr5481 forms hydrogen bond contacts with Val254 of the mobile C-terminus, so it is likely to offer additional interactions. In-house acquired NMR data, NOE cross-peaks (Chapter 5), revealed that Thr5481, Pro5482 and Gln5483 are within ~5 Å distance to Phe253, Val254, Ile255 and Pro256 of EB1c. Therefore, the tri-peptide <sup>5481</sup>TPQ<sup>5483</sup> includes important residues for the interaction with EB1c and can be used for the design and identification of molecules that can interact beyond the SxIP region.

After selection of the docking pose obtained for **1a**, and removal of the atoms that do not constitute the IP motif mimetic described in Chapter 2, Figure 4.1A, the addition of the tri-peptide described above to this anchor motif could potentially improve the chances of targeting a region outside of the hydrophobic pocket as shown by the

theoretical model in Figure 4.3A. The isoleucine side chain can be emulated by the five membered ring, whereas the methyl group occupies a similar position to the proline ring. Because both **1a** and **1d** have an amide bond after the methyl group one can make use of it to link it to the tri-peptide in a straightforward manner. Using the IP mimetic as an anchor, hopefully tighter binding will be achieved through the extra interactions with the dynamic C-terminus through contacts with the tri-peptide, as shown in the native interaction between EB1cΔ8-MACFp1.

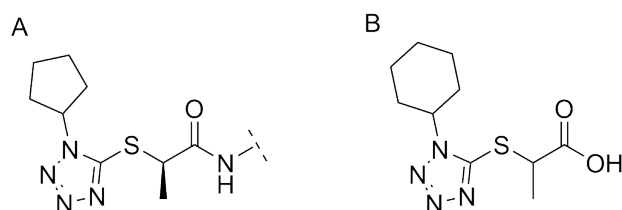


Figure 4.1 - Structures of A - 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide, IP mimetic, and B - 2-[(1-Cyclohexyl-1H-tetraazol-5-yl)sulfanyl]propanoic acid .

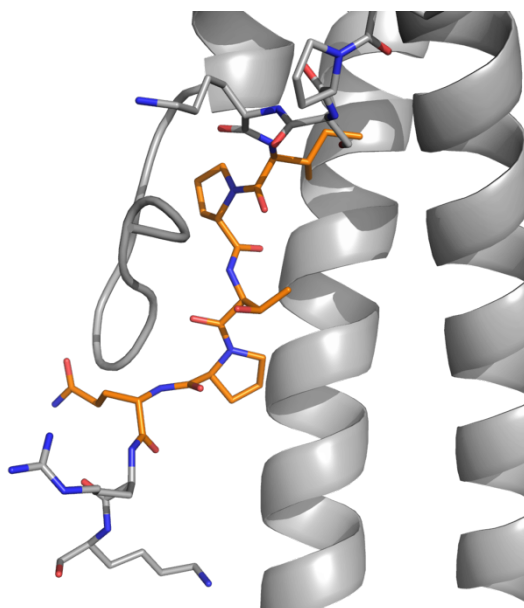


Figure 4.2 - Representation of the crystal structure of the complex EB1cΔ8-MACFp1, EB1cΔ8 is shown as grey cartoon. MACFp1 residues <sup>5475</sup>KPSK<sup>5477</sup> and <sup>5484</sup>RK<sup>5485</sup> are shown in grey. MACFp1 residues <sup>5479</sup>IPTPQ<sup>5483</sup> are shown in orange (carbon), blue (nitrogen) and red (oxygen).

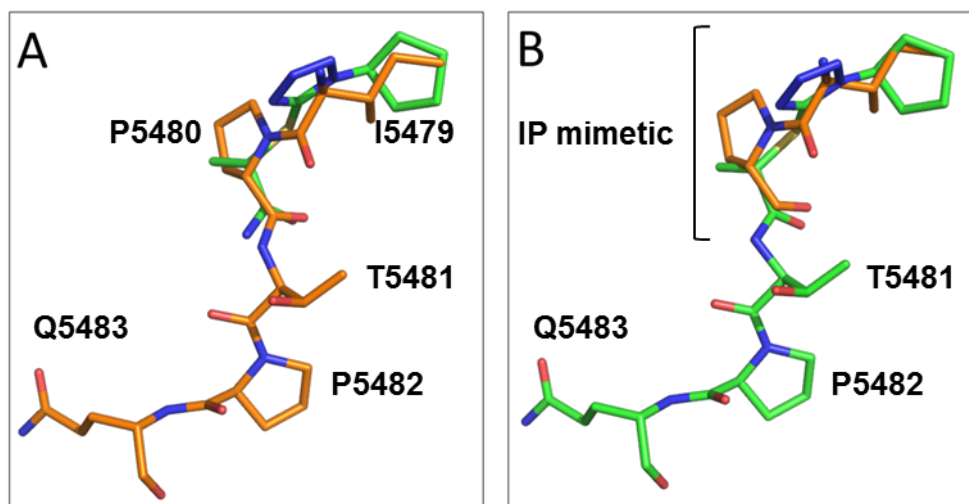


Figure 4.3 – Rational for the design of the hybrid molecule A – IPTPQ, represented in orange (carbon) blue nitrogen, and red (oxygen) sticks and IP mimetic scaffold shown in green (carbon), blue (nitrogen), red (oxygen) and yellow (sulphur). B – IP shown in orange and hybrid molecule shown in green (carbon), blue (nitrogen), red (oxygen) and yellow (sulphur).

It was, therefore, decided, to synthesise a molecule comprising the 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide scaffold and couple it with a tri-peptide sequence corresponding to Thr5481-Pro5482-Gln5483 – Figure 4.3B. To achieve this goal solid phase peptide synthesis seems to be an ideal synthesis strategy since it relies on amide bond formation and due to a developed protection/deprotection scheme allows for a reduced number of side reactions and therefore side products. Additionally, and because the compound of interest is attached to a solid support (resin), excess of reagents and side products can be washed-off.<sup>15</sup> Because this molecule encompasses a small molecule scaffold and a peptide moiety, it will be termed as hybrid molecule (2a).

Unfortunately, only a six membered ring version, 2-[(1-Cyclohexyl-1H-tetraazol-5-yl)sulfanyl]propanoic acid, of this scaffold was commercially available - Figure 4.1B. Docking results show the possibility of the six membered ring fit in the pocket, although only through biasing the docking prediction by constraint introduction. This matter will be further discussed in section 4.1.2.

#### 4.1.1 NMR based screening of the hybrid molecule – 2a

Titration of compound **2a** into EB1cΔ8, final ratio 100:1, promoted NH chemical shift changes, indicating chemical shifts of the protein backbone are perturbed upon addition of the ligand – Figure 4.4. Chemical shift changes are characteristic of a fast exchange regime, similar to what was observed for the first generation molecules.

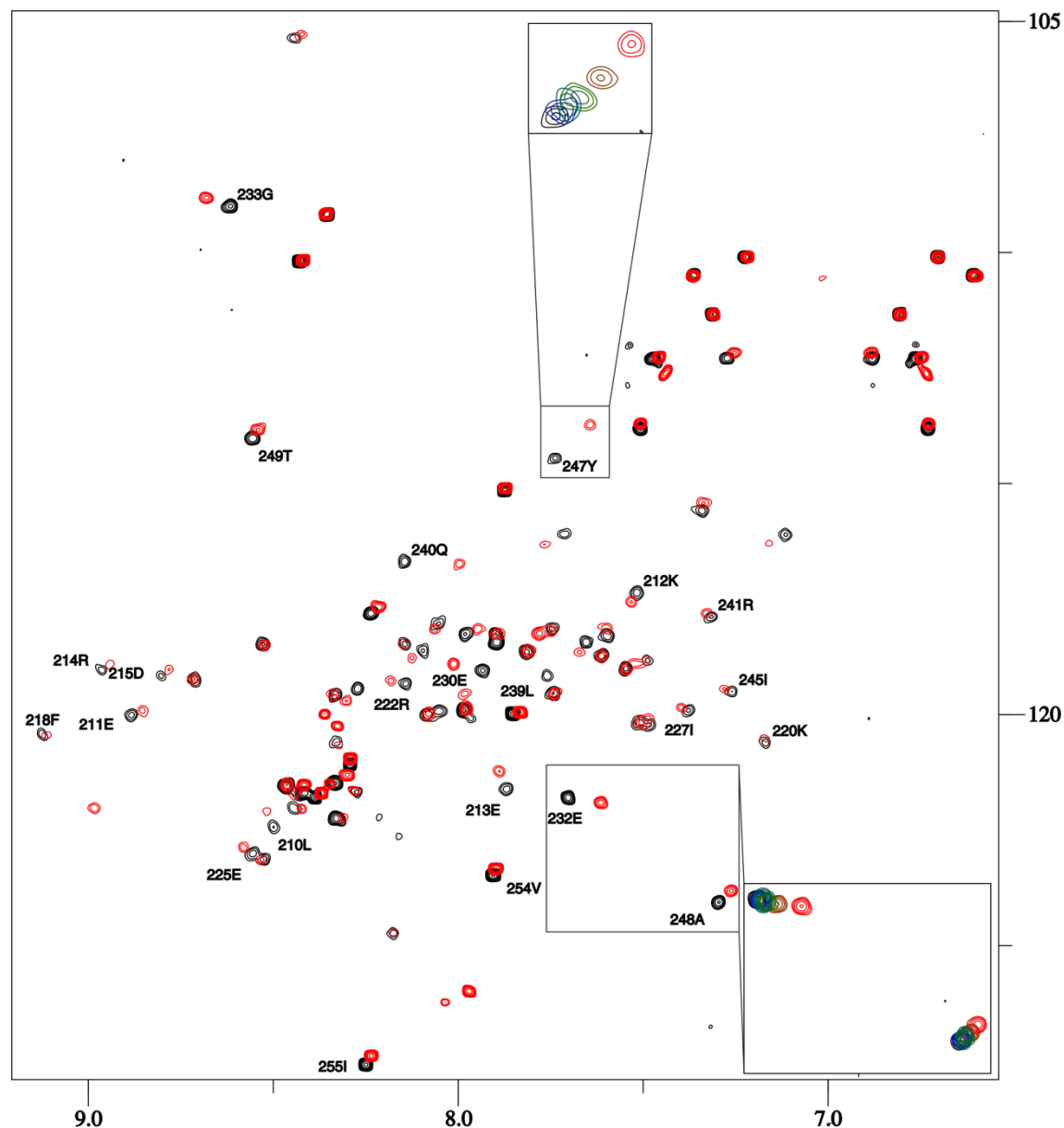


Figure 4.4 - Overlay of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1cΔ8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1c (50  $\mu\text{M}$ ) – **2a** (5000  $\mu\text{M}$ ). The insets show regions of interest and the spectra corresponding to the following **2a** concentrations, 3000 (blue), 3500 (teal), 4000 (green), 4500 (brown) and 5000  $\mu\text{M}$  (red).

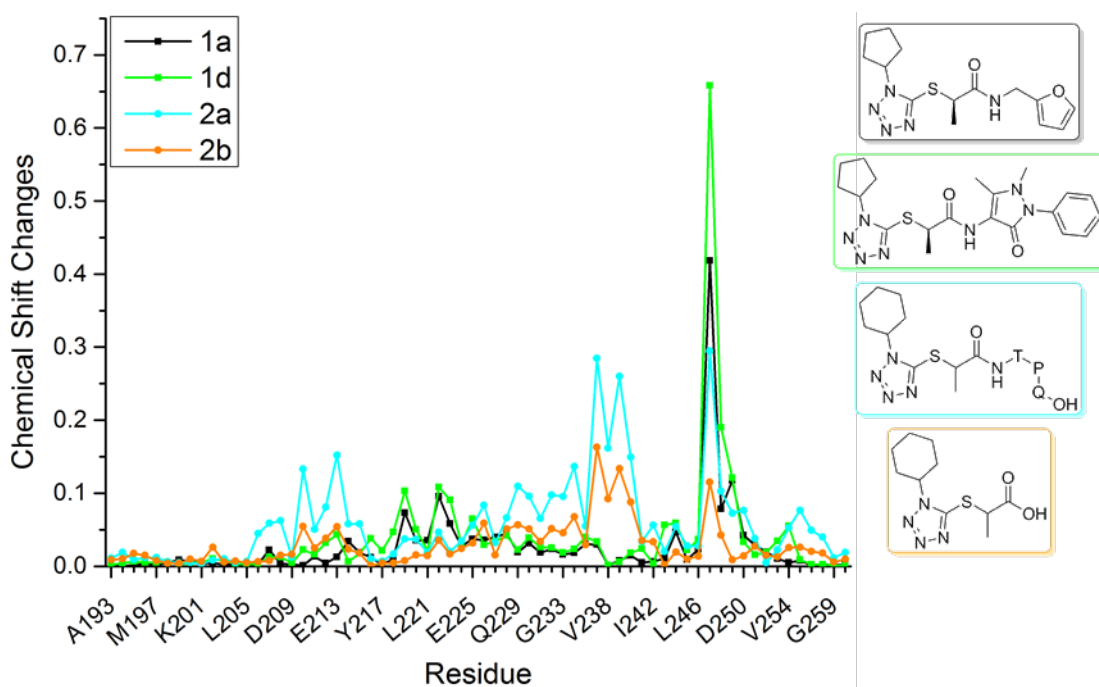


Figure 4.5 - Plot of the chemical shift perturbation distances between free EB1cΔ8 and in the presence of the respective first generation compound (**1a** (black) and **1d** (green)). Hybrid molecule is shown in cyan and 2-[(1-Cyclohexyl-1H-tetrazol-5-yl)sulfanyl]propanoic acid in orange (**2a** and **2b**, respectively).

However, when comparing the chemical shift changes promoted by the hybrid molecule, compound **2a**, with the IP mimetics identified in Chapter 2, it is clear that compound **2a** does not target the same region of EB1cΔ8 and the most prominent chemical shift changes are within Leu210-Glu213 and Gln229-Ile242 regions – Figure 4.5. The analysis of the chemical shift perturbations for the scaffold attached to the tri-peptide, molecule **2b**, reveals a similar pattern to the one observed to molecule **2a** - Figure 4.6. Titration of the tri-peptide, Thr-Pro-Gln into EB1cΔ8 does not induce any chemical shift changes into EB1's backbone resonances.

This data indicates that the specificity of the IP mimetic scaffold was lost, <sup>247</sup>YAT<sup>249</sup> and <sup>219</sup>GKLR<sup>222</sup> regions not affected by the addition of the ligand, contrarily to what happened to the first generation compounds – Chapter 2, section 2.6. Instead the chemical shift changes seem to be spread and non-specific - Figure 4.6. A possible cause can be the larger size of the ring, cyclohexyl instead of cyclopentyl, as this is the main difference between the scaffolds.

To understand why the specificity for the IP binding site was lost it was decided to perform some docking predictions.

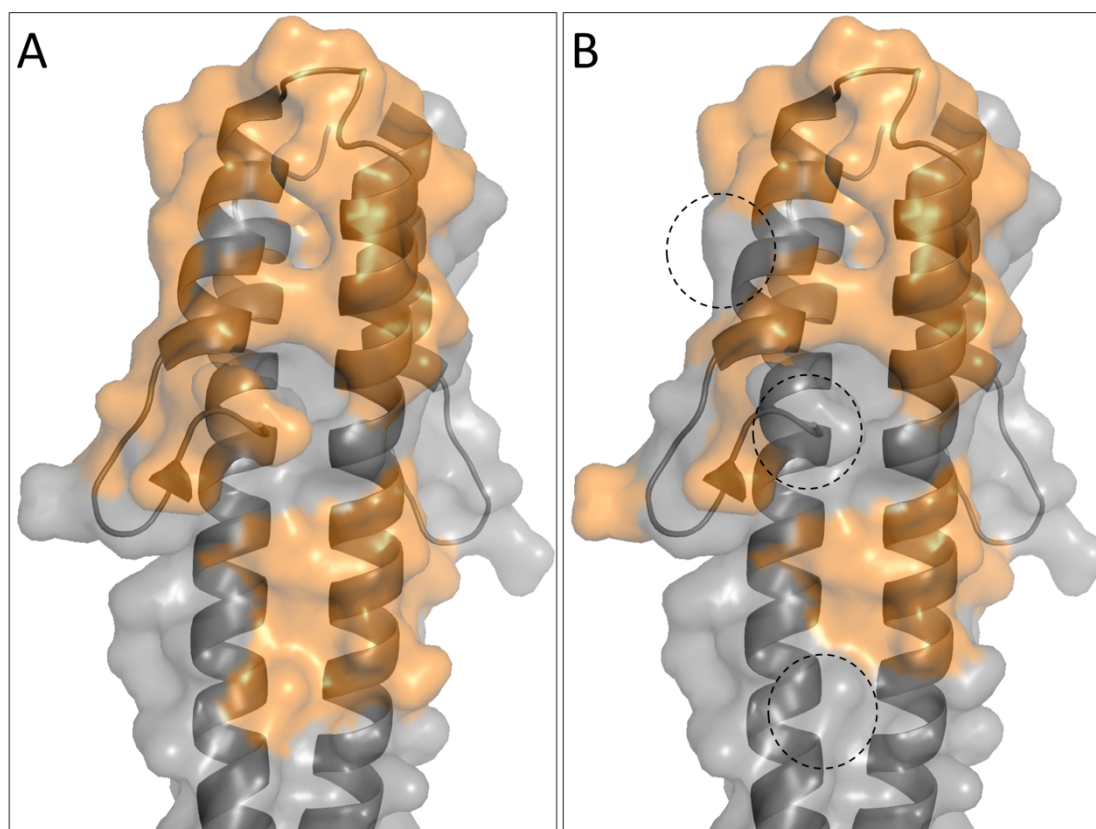


Figure 4.6 – Chemical shift perturbation mapping on the crystal structure (PDB accession code: 3GJO). Panel A corresponds to the chemical shift changes induced by the hybrid molecule (**2a**) and panel B to the changes induced by the scaffold attached to the tri-peptide, molecule (**2b**). The selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value. Dashed circles represent areas affected by titration of **2a** but not **2b**.

#### 4.1.2 Understanding of the loss of specificity to the IP binding site in the hybrid molecule

Contrarily to what was expected after the design of the hybrid molecule, the cyclohexyl ring of the IP mimetic moiety does not occupy the hydrophobic binding pocket. Instead, docking pose with higher score prediction inserts the five membered ring of proline side chain inside the hydrophobic pocket of EB1c $\Delta$ 8 – Figure 4.7A. This is not a surprise as the natural ligands reported for EB1 have invariably a proline in position 4 of the SxIP motif and structural data shows this residues side chain should occupy the hydrophobic pocket. This analysis suggests the cyclohexyl ring may be too bulky for that pocket. To prove this further, it was decided to perform another set of molecular docking predictions biasing the insertion of this ring in the pocket.



In GOLD, restraints are available for biasing searches towards dockings containing a contact between a particular protein-ligand atom pair. There is reported evidence of the benefits of using constrains in virtual screening.<sup>15></sup> The distance between a specified ligand and protein atom can be constrained to lie between minimum and maximum distance bounds. During a GOLD run, if a constrained distance is found to lie outside its bounds (*i.e.* between 1.5 and 5 Å), a spring energy term is used to reduce the fitness score:

$$E = kx^2, \text{ Equation 4.1}$$

where:

$x$  is the difference between the distance and the closest constraint bound;

$k$  is a user-defined spring constant – in this case defined as 5.

A distance constraint was defined where the distance between the six membered ring of **2a** and the Leu246 side chain of the hydrophobic pocket should be between 1.5 and 5 Å. Solutions inserting the cyclohexyl ring in the hydrophobic pocket should be in agreement with this constraint and therefore scored higher than other solutions. As expected, the ring adopts the expected position when this constraint is used, indicating that despite not being the most natural solution, the cyclohexyl ring is small enough to fit in the hydrophobic pocket of EB1c – Figure 4.7B.

However, the fact that the docking prediction only positioned the cyclohexyl ring in the pocket after the use of constraints and the first prediction re-oriented the molecule in a way that the proline ring of the tri-peptide was inserted in the pocket instead, gives support to the idea that the cyclohexyl ring is not binding in an energetically favourable way to the hydrophobic pocket of EB1c.

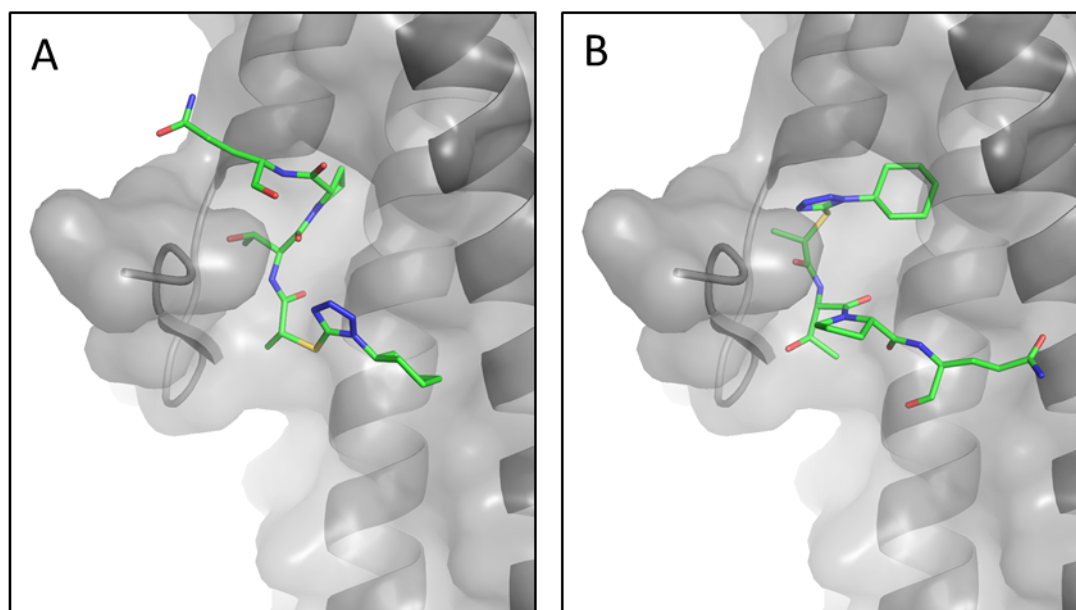


Figure 4.7 – Docking pose for compound **2a**. A- docking without any constraints. B – docking restraining the cyclohexyl ring distance to Leu246 side chain to a value between 1.5-5 Å.

## 4.2 Identification of small molecules based on the hybrid molecule

Since the in-house synthesised hybrid molecule, compound **2a**, did not yield the expected results, due to the larger size of the ring, it was thought to use the original cyclopentyl hybrid design to build a new pharmacophore model, extending the targeted region beyond the IP binding region.

### 4.2.1 Virtual screening

For this purpose, the hybrid molecule initial model, with a five membered ring was uploaded into ZINCPharmer. The uploaded structure was carefully built in order to maintain the 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide and the tri-peptide in the right spatial conformation – Figure 4.1A. When uploading the structure to ZINCPharmer, EB1c structure was also uploaded so only the points of contact between both molecules were defined as part of the pharmacophore model. Moreover, uploading the receptor (EB1c) offers the additional benefit of assigning directionality to the hydrogen bond acceptors/donors. Seven pharmacophore points were defined based on this approach: hydrophobic interaction for the cyclopentyl ring and methyl group, hydrogen bond acceptors for the CO present in the 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide scaffold and the threonine backbone and, finally, three

hydrogen bond acceptors, corresponding to the NH from the 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide, the OH of the threonine and the backbone NH of the glutamine – Figure 4.8, Table 4.1.

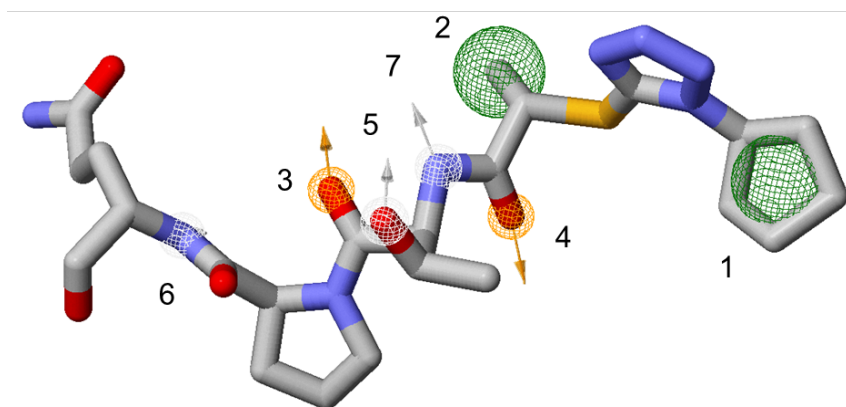


Figure 4.8 – Seven pharmacophore points defined based on the favourable conformation of 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide connected to a tri-peptide, threonine, proline and glutamine. Hydrogen bond acceptors indicated by orange arrows, hydrogen bond donors shown as light grey arrows and hydrophobic interactions represented as green spheres. The arrows represent the directionality of the hydrogen bonds defined by using the protein structure.

Table 4.1 – Summary of the pharmacophore points defined based on the hybrid molecule model and its interaction with EB1cΔ8

#	Pharmacophore Class	coordinates			radius
		x	y	z	
1	Hydrophobic	-16.08	-27.78	34.50	1
2	Hydrophobic	-18.42	-27.11	27.82	1
3	Hydrogen acceptor	-14.88	-25.70	25.15	0.5
4	Hydrogen acceptor	16.07	-25.18	28.89	0.5
5	Hydrogen donor	-12.47	-27.21	26.93	0.5
6	Hydrogen donor	-13.70	-24.47	22.22	0.5
7	Hydrogen donor	-15.27	-26.67	27.35	0.5

The pharmacophore was queried using all the contact points outlined above, but no hits were obtained. Therefore, a systematic search was performed, where all the possible combinations were searched for six, five and four pharmacophore points, in order to get the maximum number of molecules that can satisfy some of these spatial features. A total of 43388 results were yielded from this search, reducing it to 26887 molecules, after removing duplicated molecules. By applying a filter to include only molecules between 400 and 800 g.mol<sup>-1</sup>, with a maximum number of rotatable bonds of 10 and a RMSD below 0.5 Å, the number of molecules was reduced to 2429. The molecular weight range choice was based on the fact that to extend the targeted region of EB1c there is the need for larger molecules. In addition, a study performed on successful cases of inhibitor design for protein-protein interactions propose a “rule of

four” as an alternative to the Lipinski’s rule of five.<sup>53</sup> These authors propose the molecular weight threshold for small molecules that target PPI’s should be higher than 400 g.mol<sup>-1</sup> – Table 4.2.

Table 4.2 – Comparison between rule of 5, proposed as a set of *in silico* guidelines applied to drug discovery to prioritize compounds with an increased likelihood of high oral absorption and rule of 4, conceived to filter databases and accelerate the process of hit identification for protein-protein interactions modulators.

	Rule of 5 <sup>83</sup>	Rule of 4 <sup>53</sup>
Molecular weight (g.mol <sup>-1</sup> )	< 500	> 400
LogP	< 5	> 4
Number Hbond acceptors	< 10	> 4
Number Hbond donors	< 5	-
Number rings	-	> 4

The resulting 2429 molecules were then docked using GOLD and the scoring function Goldscore. As described in Chapter 2, Goldscore was the scoring function with better performance when docking the natural ligand, positioning the ligand in a similar conformation to the one shown in the crystallographic structure, 3GJO.

The importance of water molecules in docking prediction was already described in Chapter 2. GOLD provides a clever approach to deal with water molecules in docking studies. Water molecules can be allowed to rotate or switch between on and off. Adding or “toggling on” a water molecule introduces an entropic penalty to the scoring function which needs to be offset by forming hydrogen bonds to the protein and the ligand. If the hydrogen bonds formed by the water molecules do not offset the entropic penalty introduced by adding the water molecule, then the water molecule will be deselected (removed) during the genetic algorithm run.<sup>116</sup> It was decided to use this functionality offered by GOLD as it has been previously shown that there is a water molecule in EB1 binding site.<sup>31</sup> No hydrogen bond constraints were included for these docking calculations.

When using Automatic (Ligand-Dependent) Genetic Algorithm Parameter Settings in GOLD the Search Efficiency can be used to control the speed of docking and the predictive accuracy of the results, higher the efficiency more exhaustive and slower is the search. To speed up the process the *Virtual Screening* option for the GA settings in GOLD was chosen. This sets the search efficiency at 30% that can be used for

routine work and usually gives comparable predictive accuracy to the slower settings, unless the ligand has a large number of rotatable torsions.<sup>16></sup>

The ten best solutions were saved and visually inspected. The criterion used for visual inspection was that the pose included interactions outside the hydrophobic pocket, extending to a lower region occupied by the tri-peptide Thr-Pro-Gln. Through this process the number of molecules to be further studied was reduced to 106 molecules – Figure 4.9.

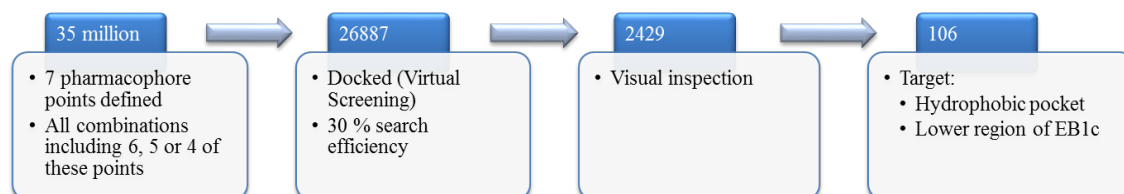


Figure 4.9 – Process for selection of hit compounds that can target EB1c using as initial scaffold the IP motif mimetic attached to a tri-peptide.

#### 4.2.2 Selection of candidate molecules

The 106 molecules selected based on their binding pose were re-docked using a search efficiency of 200% and the 100 best solutions were saved into the result file for further refinement/analysis.

Up to this stage no filters were used to remove molecules that reveal potential problems in terms of ligandability. The rule of 4 used previously relates only with the capacity of being a protein-protein interaction modulator. The use of filters or alerts to remove or flag compounds that may not be suitable to become a drug, has been widely used by pharmaceutical companies.<sup>16></sup> Therefore it was decided to use an online tool that could flag potential undesirable compounds before moving forward into candidate selection.

The SmartsFilter (<http://pasilla.health.unm.edu/tomcat/biocomp/smartsfilter>)<sup>16></sup> uses a wide set of filters that can be easily applied to a subset of candidate molecules. There are seven different filters, from which Blake<sup>162</sup>, Glaxo<sup>163</sup>, ALARM NMR<sup>164</sup>, Oprea<sup>165</sup> and PAINS – Pan-Assay Interference compounds<sup>166</sup> were chosen to avoid molecules with undesirable characteristics at an early stage of the drug discovery process.

The Blake filter<sup>16></sup>, is based on the work of James Blake, and concerns a model of desirable properties for good oral bioavailability.

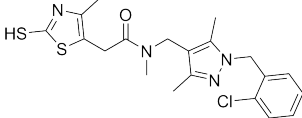
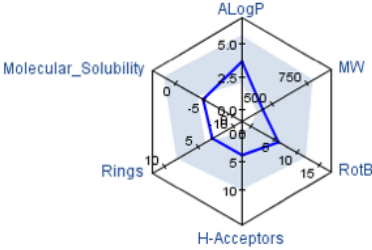
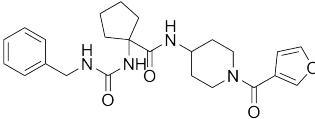
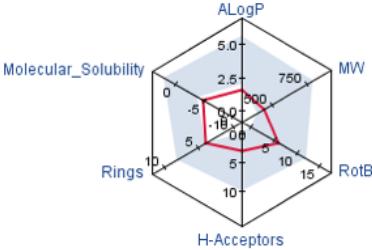
The Glaxo filter<sup>164</sup> has three different subfilters, where only two were used to evaluate our subset of molecules, namely “unsuitable leads” and “reactive”. Whereas the first is clearly described by the name, the latter is based on the existence of reactive species that cannot be pooled together in HTS study. At this stage it was not known the number of candidate molecules to be screened and the pooling of compounds when performing HTS is commonly used to save time and money.

Because solution-state NMR is the primarily chosen method for compound screening in this project, the filter ALARM NMR<sup>165</sup> was also applied. This filter can be applied to identify potential false positive molecules in NMR based screenings, by reacting with protein’s thiol groups. The Oprea filter, developed by Tudor Oprea<sup>165</sup>, is based on a multi-objective analysis of drug-like properties. Finally, the PAINS filter<sup>166</sup>, developed by Jonathan Baell and Georgina Holloway, proposes a method to identify frequent hitters, promiscuous compounds, widely found in HTS assays. Compounds that did not successfully pass these filters were not removed. Instead, they were marked as “pass” or “fail” for each filter and this information used in the final assessment of suitable candidates to be screened.

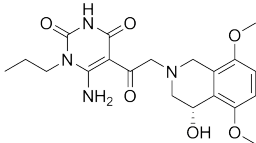
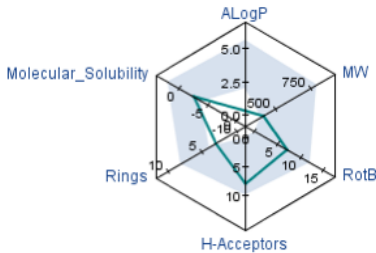
At this stage, there was sufficient knowledge about these compounds that would permit to make a decision of which should be tested. A multi-objective analysis was performed, assessing two parameters, the average Goldscore value for the 100 obtained poses and solubility. In contrast to what was done before (Chapter 2) compounds with solubility lower than - 4 were still included for the final analysis. Three compounds were chosen based on their Pareto Rank position, availability for purchase and molecular diversity - Table 4.3. The method chosen to select these three candidate molecules has been summarised in Figure 4.9 and Figure 4.10.

ZINC40017212 was the fifth best compound in terms of the multi-objective analysis. The average Goldscore for 100 poses was high, 66, with a top score of 75. It failed for three of the filters, unsuitable leads, ALARM NMR and PAINS, which can be attributed in part to its low solubility value, - 4.918. The pharmacophore model of this molecule includes the hydrophobic contacts of the cyclopentyl ring and the methyl group, in addition to a possible hydrogen bond acceptor (corresponding to the carbonyl of 2-((1-cyclopentyl-1H-tetrazol-5-yl)thio)propanamide) and a hydrogen bond donor (corresponding to the OH of Thr), respectively pharmacophore points 1, 2, 4 and 5 – Table 4.1; Figure 4.11A.

Table 4.3 – Selected molecules for testing from the pharmacophore model built from the hybrid molecule. Final rank indicates the overall position of each molecule after a Pareto ranking analysis of the 100 solutions for the 106 molecules. Log S refers to the logarithm of aqueous solubility value calculated by Pipeline Pilot<sup>16</sup>. Unsuitable leads, reactive, ALARM NMR, Blake and Oprea columns state whether a compound passed (green symbol) or failed (red symbol) each filter. Finally, the molecular properties radar plot includes six molecular properties and the blue shaded area represents the chemical space defined for PPI modulators.<sup>53</sup>

Molecule name	Final Rank	Log S	Unsuitable leads	Reactive	ALARM NMR	Blake	Oprea	PAINS	Structure	Molecular properties
ZINC40017212	5	-4.9	✗	✓	✗	✓	✓	✗		
ZINC31040053	71	-4.9	✓	✓	✗	✓	✓	✗		

Design and identification of inhibitors beyond SxIP motif

Molecule name	Final Rank	Log S	Unsuitable leads	Reactive	ALARM NMR	Blake	Oprea	PAINS	Structure	Molecular properties
ZINC12929029	101	-2.5	✓	✓	✗	✗	✗	✓		

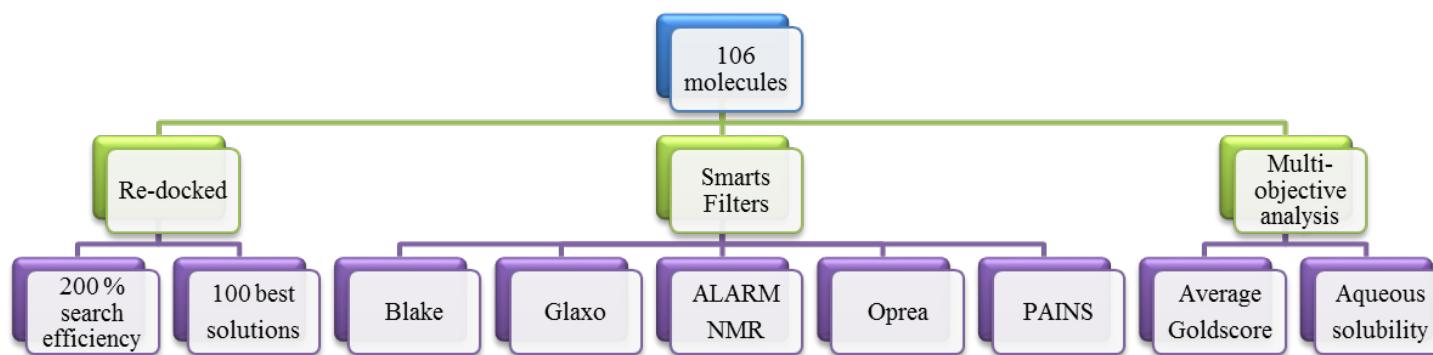


Figure 4.10 – Process summarising the final selection of hit compounds to target EB1c based on the scaffold the IP motif mimetic attached to a tri-peptide.



In terms of docking pose, ZINC40017212 seems to interact through hydrophobic contacts with the hydrophobic pocket of EB1c, Lys220, Arg222, Leu221, Glu225, Tyr247, Ala248 and Thr249, but also with a lower region of EB1c, including the aromatic patch, Phe216-Tyr-Phe218 and the C-terminal tail, Phe253, Val254 and Pro256 – Figure 4.12, A and B.

ZINC31040053 presents a more modest average Goldscore value of 53 with a top score of 59. In terms of filters, fails for both ALARM NMR and PAINS. Again, its solubility value is quite low, – 4.911, and that can be a disadvantage for solution NMR screening techniques. In terms of pharmacophore model, this molecule coincides with pharmacophore points 1, 2, 3 and 5 – Table 4.1; Figure 4.11B. The best scored binding pose shows the existence of hydrophobic interactions within the hydrophobic pocket and aromatic patch, as well as hydrophobic interactions with C-terminus residues such as Val254, Ile255 and Pro256. The hydrophobic pocket seems to be almost invariably occupied by the phenyl ring, especially for the top scored solutions - Figure 4.12, panels C and D.

ZINC12929029 shows the lowest Goldscore average value 45 with a top score of 55. The pharmacophore points emulated by this molecule are the same as for ZINC40017212, 1, 2, 4 and 5 - Table 4.1; Figure 4.11C. The genetic algorithm poses the tetrahydroisoquinoline moiety inside the hydrophobic pocket of EB1c. It is not clear whether this scaffold is not too bulky for the hydrophobic pocket of EB1c as previously reported for the cyclohexyl ring. The fact that this moiety, despite its larger size, is more planar and less flexible than the cyclohexyl ring may justify the fit inside the pocket. In addition, this molecule seems to bring extra interactions, namely two hydrogen bonds. One of the carbonyls act as hydrogen bond acceptor from the backbone NH of Val 254 at the same time the NH next to it functions as hydrogen bond donor to the carbonyl oxygen of Gly252 – Figure 4.12, panels E and F.

These three molecules were therefore purchased and subjected to NMR screening, through  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra analysis upon titration of each compound into EB1c $\Delta$ 8. For clarity the molecules will be referred to as **3a** (ZINC40017212), **3b** (ZINC31040053) and **3c** (ZINC12929029).

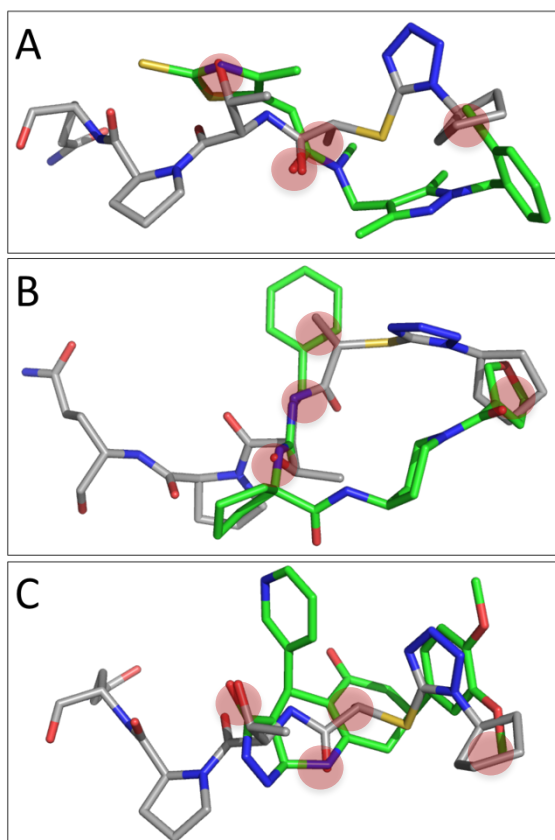


Figure 4.11 – Molecules selected for screening based on the pharmacophore model created for the hybrid molecule. A – ZINC40017212, B – ZINC31040053 and C – ZINC12929029, represented as green sticks, superimposed with the ideal 3D structure designed for the hybrid molecule, showed as grey sticks. Areas highlighted in red correspond to the pharmacophore contacts emulated by these molecules.

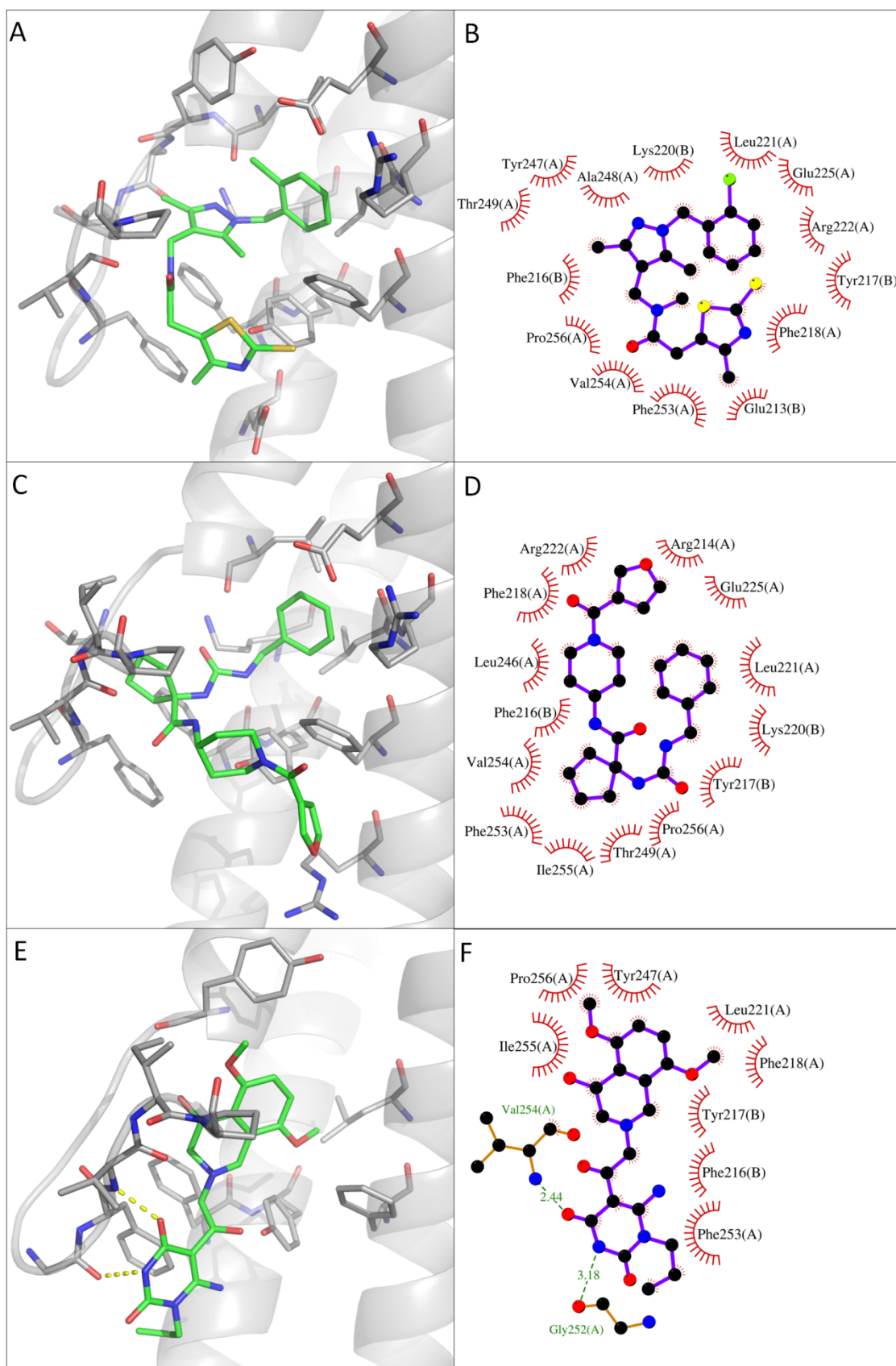


Figure 4.12 – Top scored binding pose, left hand side panel, A- ZINC40017212 (3a), C – ZINC31040053 (3b) and E – ZINC12929029 (3c). 2D representation of the interactions predicted from the docking between B - ZINC40017212 (3a), D – ZINC31040053 (3b) and F – ZINC12929029 (3c).

#### 4.2.3 NMR based screening for hit candidates based on the hybrid molecule

NMR titrations were, once again, the chosen method to screen the three candidate molecules obtained from ligand-based design approach based on the hybrid molecule. The results indicate that only **3a** and **3b** promote chemical shift perturbations in the EB1cΔ8 backbone (NH).

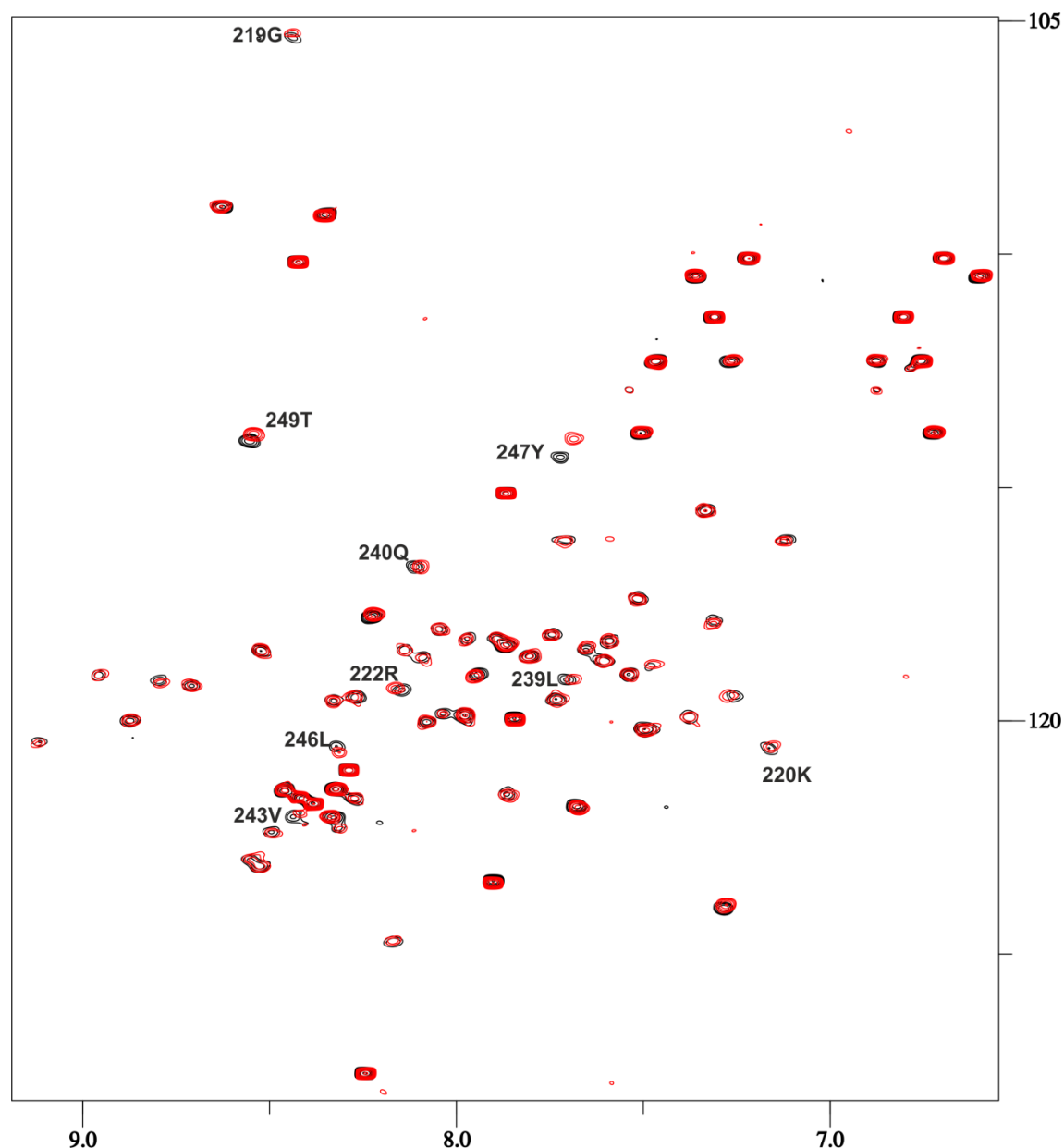


Figure 4.13 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1cΔ8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1cΔ8 (50  $\mu\text{M}$ ) – **3a** (1000  $\mu\text{M}$ ). Both samples contained 4% DMSO- $d_6$ , to ensure solubility of **3a** and for consistency (in the free protein sample).

The addition of compound **3a** - affects the chemical shifts of amino acids recognised as part of the EB1cΔ8 hydrophobic pocket – Leu246, Tyr247 and Thr249, being a

good indication that this compound, contrarily of what was observed with the six membered ring hybrid molecule, targets the desired region of EB1cΔ8 – Figure 4.13. Unfortunately, due to its poor aqueous solubility **3a** was only tested up to a concentration of 1 mM. The final protein:ligand ratio obtained was 1:20, different from the ratios used for the first generation compounds (1:100). Therefore, the chemical shift changes observed are not as large as the ones observed for **1a** and **1d** – Chapter 2.

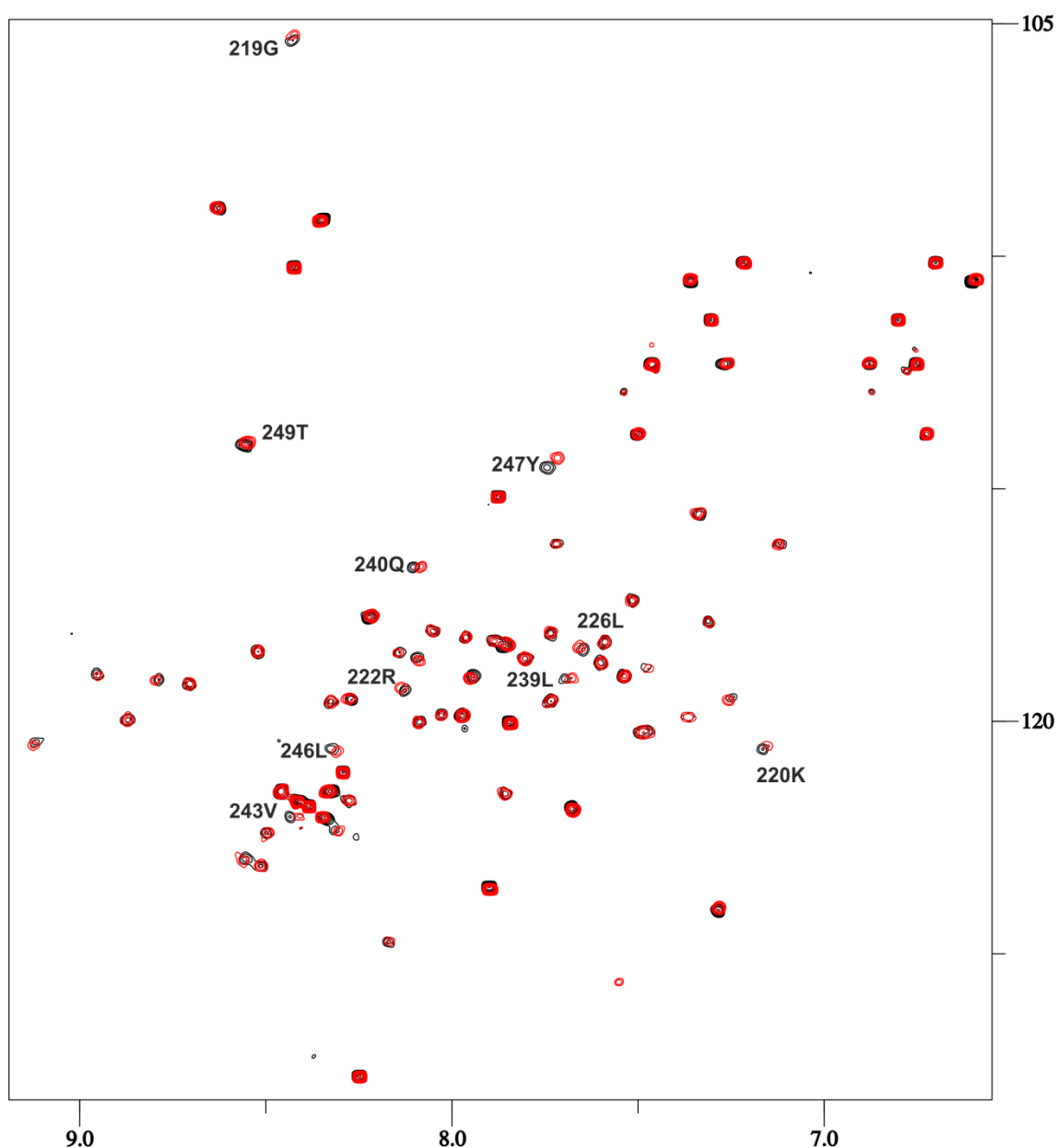


Figure 4.14 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1cΔ8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1cΔ8 (50  $\mu\text{M}$ ) – **3b** (1000  $\mu\text{M}$ ). Both samples contained 2% DMSO- $d_6$ , to ensure solubility of **3b** and for consistency (in the free protein sample).

Compound **3b** promotes similar chemical shift changes in terms of affected amino acids and magnitude to the ones observed to **3a**. These are mainly confined to the hydrophobic binding site of EB1cΔ8 – Figure 4.14. Similarly to the other compound, **3a**, due its poor aqueous solubility the final concentration was 1 mM, whereas the final concentration reached for the first generation molecules (Chapter 2) was 5 mM. Compounds **3a** and **3b**, with predicted aqueous solubility values of - 4.918 and - 4.911, respectively, interact with EB1cΔ8 but their aqueous solubility is a limitation to obtain a  $K_d$  value by NMR or even to study further through NOE experiments as it was done for compound **1a**.

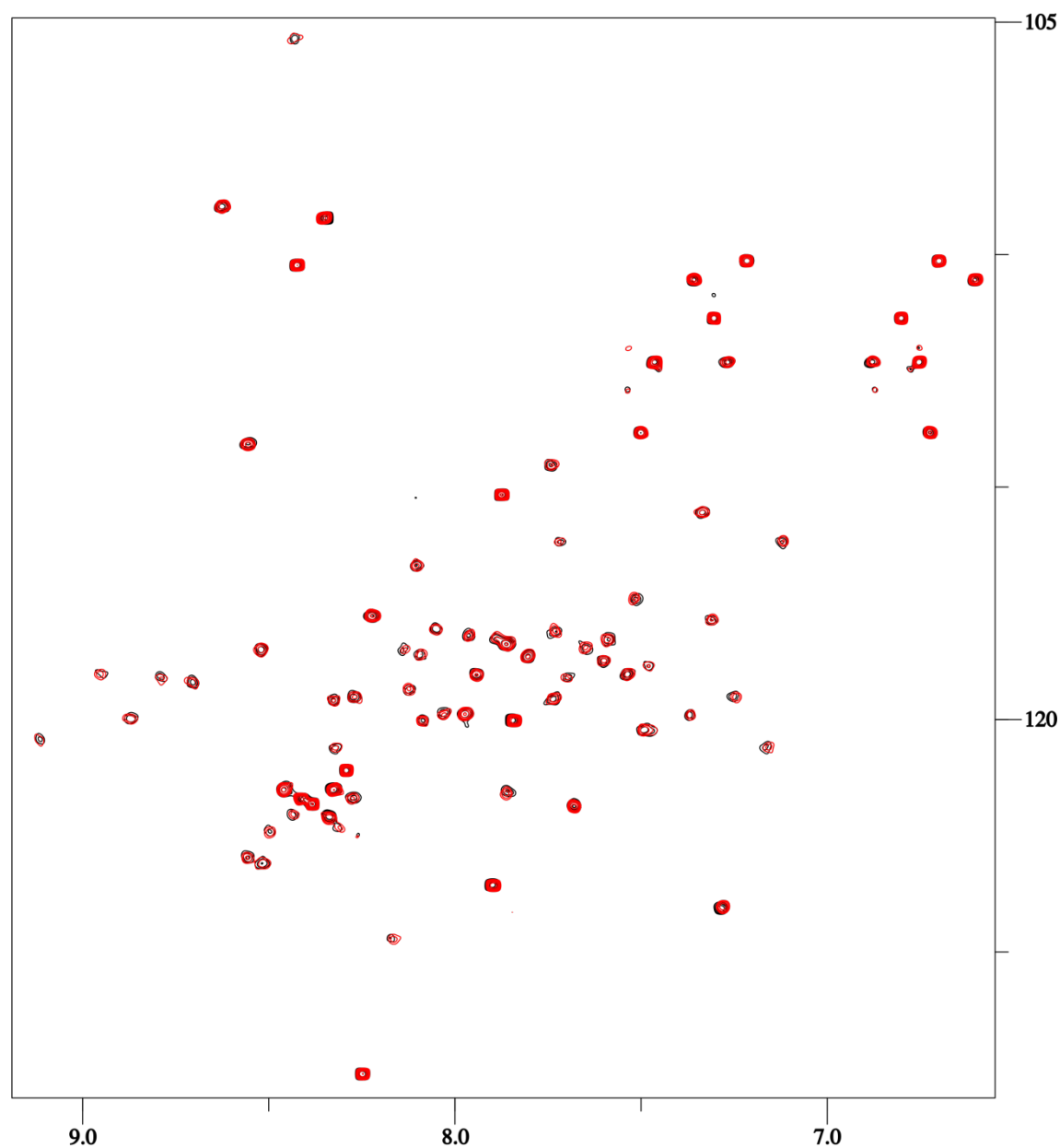


Figure 4.15 - Overlay of  $^1\text{H},^{15}\text{N}$ -HSQC spectra recorded at 600 MHz – in black is presented EB1cΔ8 in the free form (50  $\mu\text{M}$ ) and in red the complex EB1cΔ8 (50  $\mu\text{M}$ ) – **3c** (1000  $\mu\text{M}$ ). Both samples contained

2% DMSO-d<sub>6</sub>, to ensure solubility of **3c** and for consistency (in the free protein sample).

Finally, for compound **3c**, no chemical shift changes were observed (Figure 4.15) at a final concentration of 1 mM, and despite this compound having a larger aqueous solubility predicted value (calculated logS = - 2.524), it was decided to not proceed with the titration since chemical shift changes were not observed at this concentration, it would be unlikely that they would be observed at higher concentrations.

In terms of the distribution of the chemical shift perturbations for **3a** and **3b**, they follow the same pattern as observed for the molecules identified as IP mimetics – first generation compounds. Therefore the IP mimetic activity was regained. Analysis of the docking poses for these compounds, Figure 4.12, reveals the hydrophobic pocket is invariably occupied by an aromatic ring, benzyl for compound **3a** and 2-chloro-benzyl for compound **3b**. Despite its size, same number of carbons as the cyclohexyl group, being planar makes benzyl a tolerated chemical moiety for the hydrophobic pocket of EB1c, whereas the cyclohexyl, with additional hydrogens and a flexible tri-dimensional conformation, is not. Moreover, one can observe the appearance of subtle chemical shift change for the C-terminus region Glu251-Val254 when comparing with the best compound resultant from the first generation inhibitors – Figure 4.17.

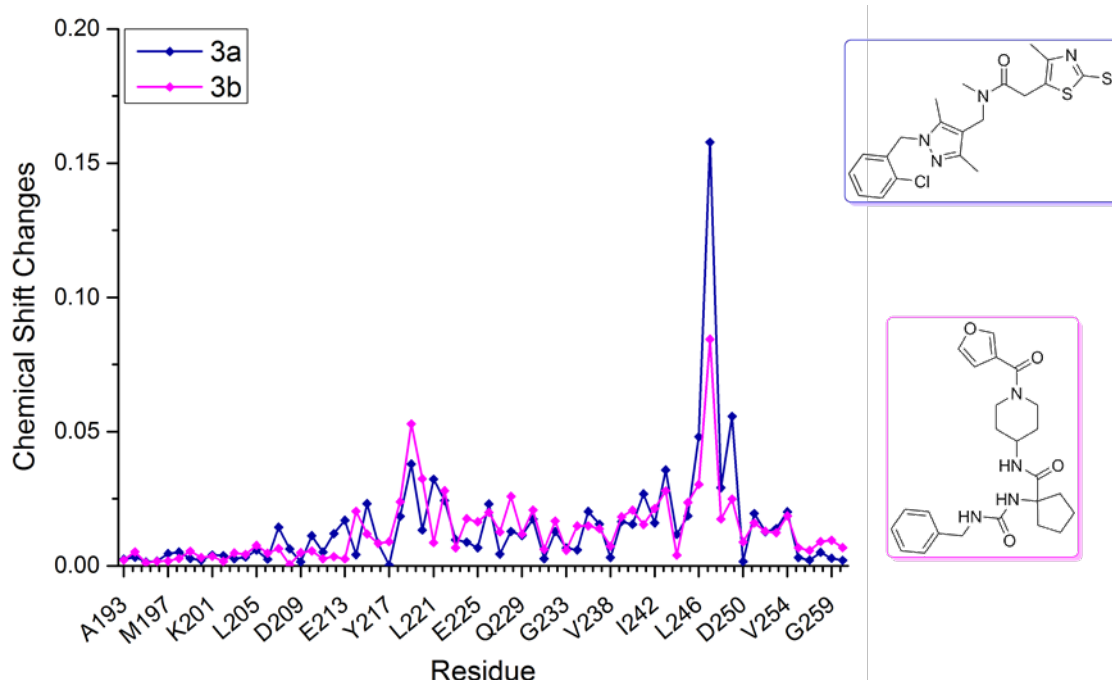


Figure 4.16 - Plot of the chemical shift perturbation sum differences between free EB1cΔ8 and in the presence of the respective second generation compound (**3a** and **3b**, respectively).

In summary, the second generation compounds, based on the hybrid molecule scaffold

successfully target EB1c, not only the hydrophobic binding site – IP mimetic activity, as well the C-terminus, to a smaller extent. Despite their solubility proved to be a major limitation, the success of this approach gives extra support to the molecular modelling approach used in this project for the identification of small molecules that can bind to a protein-protein interface.

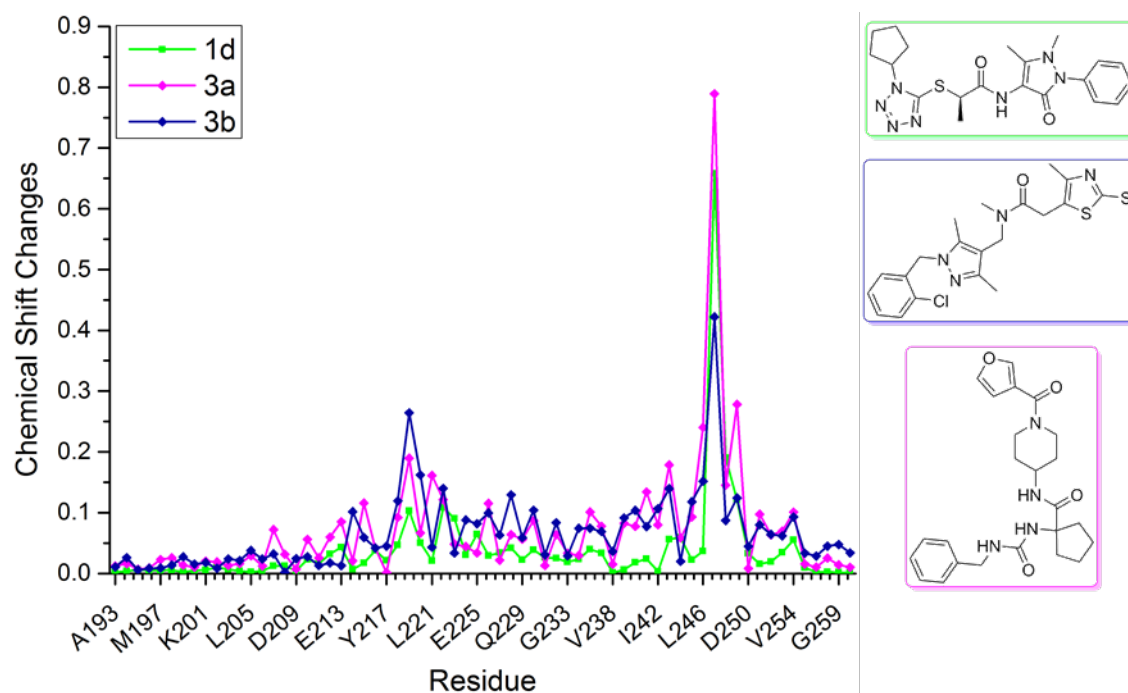


Figure 4.17 - Plot of the chemical shift perturbation sum differences between free EB1cΔ8 and in the presence of the respective first and second generation compounds (**1d**, **3a** and **3b**, respectively). The chemical shift changes for **3a** and **3b** were normalised by a factor of five since the final concentration of these compounds was five times smaller than the final concentration for compound **1a**.

The hybrid molecule design, due to its novelty, can be considered as another type of approach to target protein-protein interactions. The addition of a peptide scaffold to an anchor scaffold based on an organic molecule may bring many advantages for the design of protein-protein modulators. Through this method, it is possible to target a larger area without loss of aqueous solubility due to the peptide moiety. The shortcomings of using a peptide scaffold are chemical instability, being prone to hydrolysis and oxidation, tendency for aggregation, short half-life and fast elimination, poor oral bioavailability and low membrane permeability.<sup>167</sup> Multifunctional peptides can be used to overcome some of these disadvantages.<sup>168</sup> However, that is outside of the scope of this project.

In the next chapter, a detailed study of the interaction of EB1 with SxIP motif



containing peptides is described, since it is important to understand the contributions of each region for the interaction and this understanding can be used for the improvement of the design/identification of small molecules that can interact with EB1.



## Chapter 5. BIOPHYSICAL CHARACTERISATION OF PHYSIOLOGICAL EB1 INTERACTIONS INVOLVING MACF2

From the previous data, Chapter 2, it is known that the SxIP motif provides specificity for EB1c targeting but it is not sufficient to achieve a high affinity interaction. In Chapter 4, attempts to extend the interactions beyond the SxIP motif were made with limited success.

The aim of this chapter will be to investigate how does EB1c interact with a native ligand – MACF2(Uniprot<sup>47</sup> accession number - Q03001), and how does each region of the peptide and the protein contribute into the binding affinity. The contribution of different ligand regions to the binding to EB1c was evaluated using three peptides containing the SxIP motif, SKIP (4MACF), SKIPTP (6MACF) and KPSKIPTPQRK (11MACF), respectively. The role of the C-terminal flexible tail of the protein was also studied by using two different EB1 C-terminal domain constructs with and without this region – EB1c $\Delta$ 8 (191-260) and EB1c $\Delta$ 16 (191-252). With this information new structural information will be provided and can be used for the design of new EB1 targeting scaffolds.

Finally, higher affinity sequences based on the native sequence were designed through the use of mutants. These higher affinity sequences provide interesting insights on the binding mechanism to EB1c and provide guidance for future modulator/probe design.

### 5.1 NMR screening of MACF2 peptides against EB1c $\Delta$ 8

Chemical shift perturbation (CSP) is the only technique that can directly provide an estimate of  $K_d$  values and location of the binding site from the same set of measurements<sup>94</sup>, hence, it was used to characterise the interaction of MACF2 based peptides with EB1c $\Delta$ 8. Initial screening revealed that interactions between EB1c and 4MACF and 6MACF are weaker than for the 11MACF. For that reason, the concentrations used for the titrations of the shorter ligands were 250, 500, 1000, 2500 and 5000  $\mu$ M, respectively 5, 10, 20, 50 and 100-fold excess to the protein. Resonances of both complexes with 4MACF and 6MACF are, generally, in fast exchange and the spectra presents an overall good dispersion of peaks it was possible to assign all

resonances just by following the peak changes from free protein.

On the other hand, 11MACF seemed to have higher affinity for EB1c and therefore the used concentrations are much smaller, 12.5, 25, 75, 82.5, 100, 125, 150, 200 and 400  $\mu\text{M}$ .

4MACF, corresponding to the SxIP motif, promotes relatively small, but specific chemical shift changes – Figure 5.1.

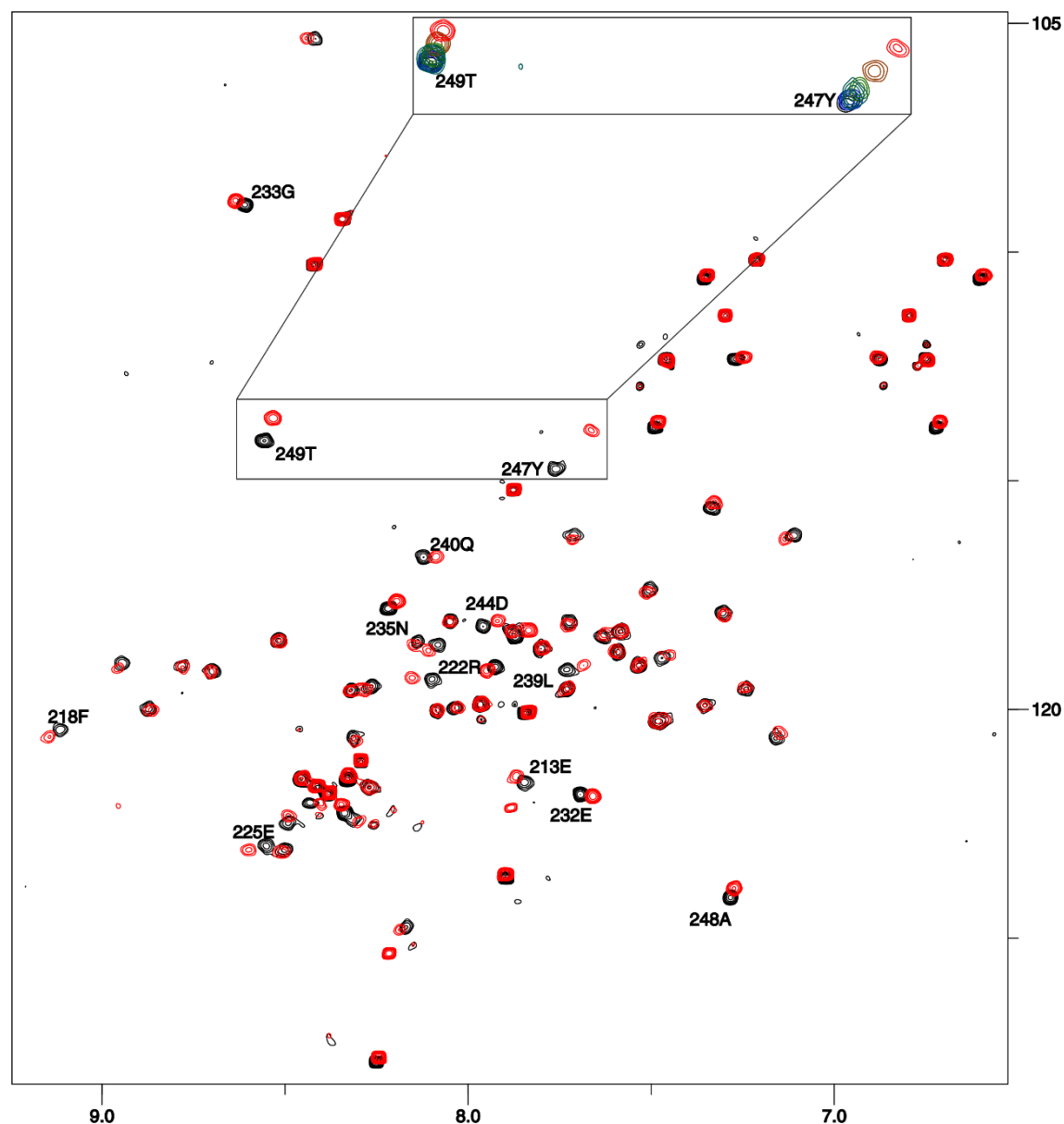


Figure 5.1 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 8 (50  $\mu\text{M}$ ) recorded at 600 MHz – in the free form (black) and in the presence of 4MACF (5000  $\mu\text{M}$ , red). The insets show regions of interest and the spectra corresponding to the following 4MACF concentrations, 250, 500, 1000, 2500 and 5000  $\mu\text{M}$ .

All chemical shift changes for 4MACF titration correspond to a fast exchange regime,

where a progressive change in peak position is observed at all ligand concentrations used without increase in line-width, indicating weak interaction. The observed chemical shift is therefore a weighted average of the chemical shifts for the free protein and the complex until the binding partner has been added to excess.<sup>94</sup>

The 6MACF addition has much larger effect on the EB1c signals, compared to the shorter 4MACF peptide (Figure 5.2).

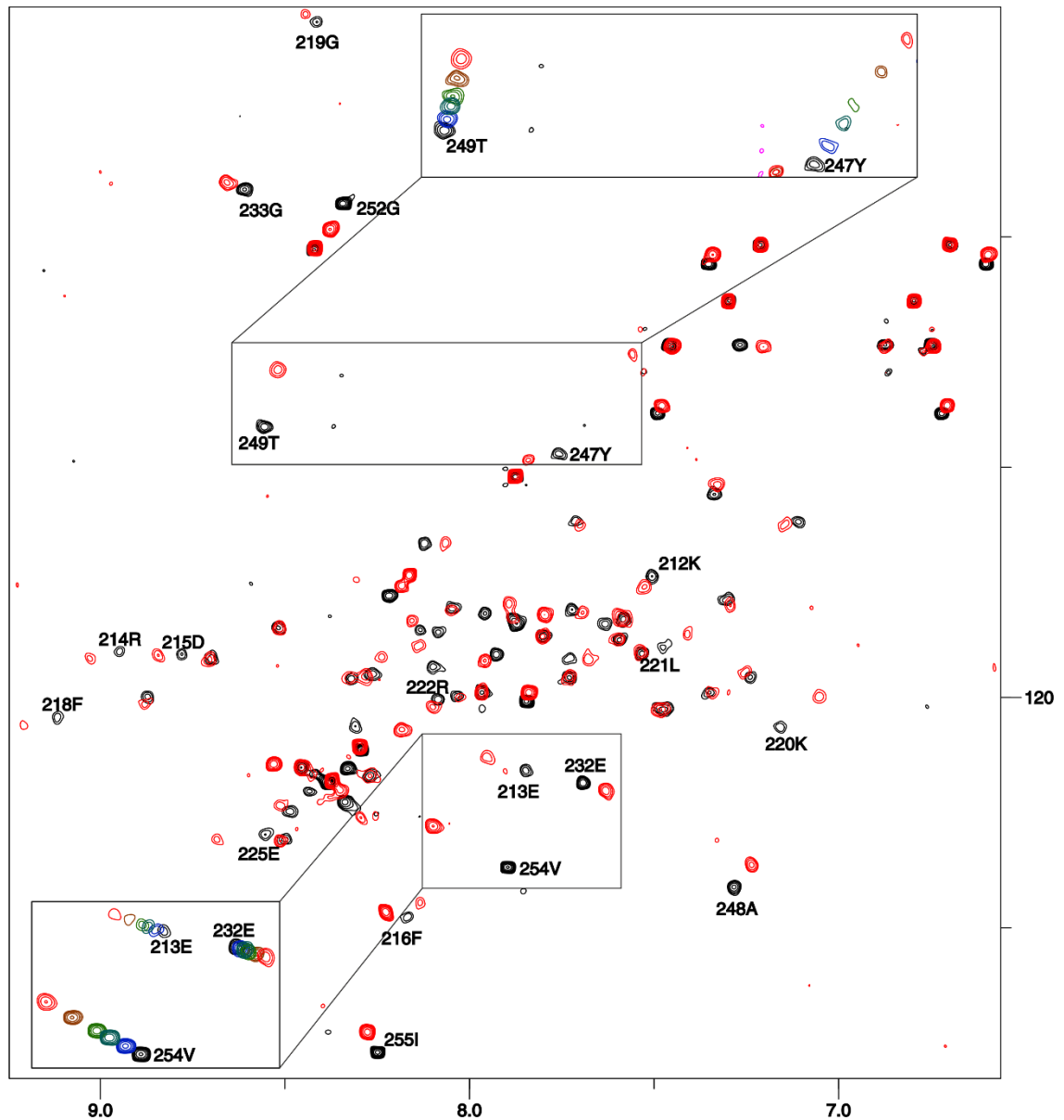


Figure 5.2 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 8 (50  $\mu\text{M}$ ) recorded at 600 MHz – in the free form (black) and in the presence of 6MACF (5000  $\mu\text{M}$ ). The insets show regions of interest and the spectra corresponding to the following 6MACF concentrations, 250, 500, 1000, 2500 and 5000  $\mu\text{M}$ .

To the same ligand excess, the chemical shift changes are overall larger and affect a

larger number of residues that were not affected before. Tyr247 ( $\Delta\delta = 0.86$  ppm) is again the most affected residue with larger chemical shift change, and it is noteworthy the peak broadening at a 20-fold excess of the ligand (1000  $\mu\text{M}$ ). Glu213 also shows line width broadening, but subtler than the one observed for Tyr247. Peak broadening suggests this complex is closer to an intermediate exchange regime. Line broadening and the larger chemical shift changes are good indicators of stronger binding when compared with 4MACF.

The addition of 11MACF has much larger effect on the EB1c $\Delta$ 8 spectra than the shorter peptides (compare Figure 5.1, Figure 5.2 and Figure 5.3).

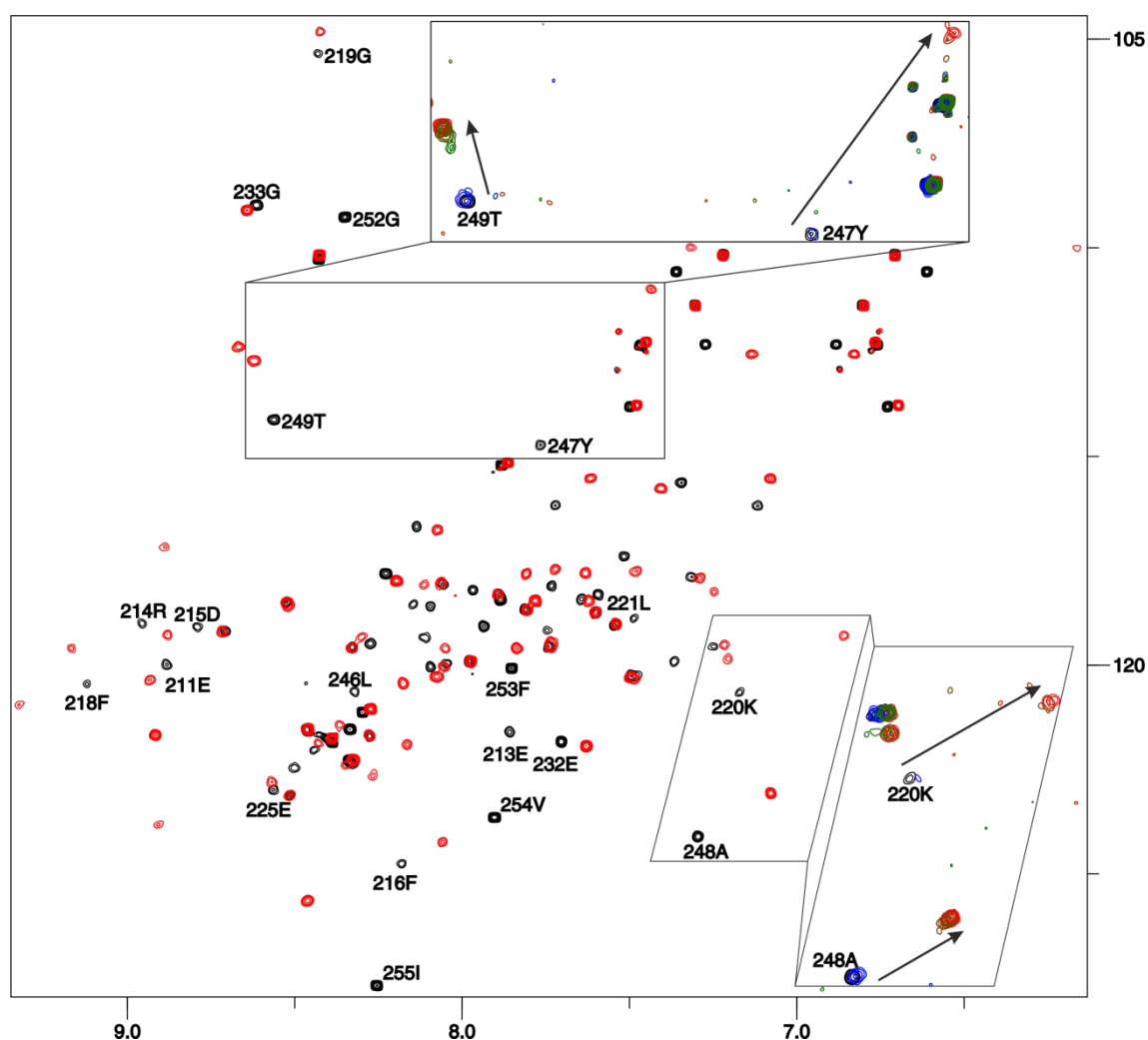


Figure 5.3 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 8 (100  $\mu\text{M}$ ) recorded at 600 MHz in the free form (black) and in the presence of 11MACF (400  $\mu\text{M}$ , red). The insets show regions of interest and the spectra corresponding to the following 11MACF concentrations, 12.5, 25, 75, 82.5, 100, 125, 150, 200 and 400  $\mu\text{M}$ .

At the 11MACF concentration of 12.5  $\mu\text{M}$  (1:0.125 protein:peptide ratio) a large

number of resonances are severely broadened with a minimal change in chemical shifts (Figure 5.3, inserts). Despite good peak dispersion, due to the accentuated peak broadening it was not possible to simply follow the chemical shift changes. Therefore, the NH resonances of the complex were assigned using triple resonance experiments similarly to what was performed for free EB1c. Further increase in the peptide concentration leads to additional broadening and then disappearance of the cross-peaks above 25  $\mu\text{M}$  (1:0.25 protein:peptide ratio). The cross-peaks reappear at concentrations above 100  $\mu\text{M}$  (1:1 protein:peptide ratio) at new positions. These changes were observed for the chemical shift differences between the free and the bound state above 0.45 ppm; resonances of Lys220 ( $\Delta\delta = 0.61$  ppm), Tyr247 ( $\Delta\delta = 1.48$  ppm), Ala248 ( $\Delta\delta = 0.45$  ppm), Thr249 ( $\Delta\delta = 0.54$  ppm) are shown as clear examples of such changes (Figure 5.3, insets). The described spectral changes correspond to the intermediate exchange regime.<sup>93</sup> The majority of other resonances showed progressing shift changes on the peptide addition, often accompanied by a limited broadening of cross-peaks; this was observed for the chemical shift differences between the free and the bound state below 0.45 ppm. Above 1:1 protein:peptide ratio chemical shift changes were very limited, showing that nearly all protein was in the bound state; in contrast, significant spectral changes were observed for shorter peptides at the ratio as high as 1:100. Overall the resonances are in an intermediate exchange regime where extensive line broadening occurs accompanied by a progressive chemical shift change. The intermediate exchange and degree of the chemical shift changes demonstrate much higher affinity of 11MACF for EB1c $\Delta$ 8 when compared with the shorter peptides.

In order to slow the exchange rate between free EB1c $\Delta$ 8 and when bound it was decided to perform another titration experiment at lower temperature, 10°C. The following protein-ligand ratios were used: 1:0, 1:0.5, 1:0.75, 1:1 and 1:1.5 – Figure 5.4 and Figure 5.5. For Phe218 it is very clear the coexistence of two intermediate states in solution at the 1:0.5 ratio. The third column in Figure 5.5 (top panel) shows the existence of three peaks – one corresponds to free EB1, followed by two peaks of similar intensity, these correspond to a ligand molecule bound to only one of the symmetrical binding sites. Finally, it is still possible to see the existence of a smaller intensity peak corresponding to the fully saturated bound form, where the two binding sites are occupied by the ligand. Despite exhibiting quite a small chemical shift

perturbation upon ligand binding, Gly219, is also an interesting case where it can be seen the free protein peak (black, second column), splitting in two peaks with similar intensity (blue, third column), representing the two complexes with only one ligand bound. With more ligand added (between 1:0.75 and 1:1 ratio), the peak evolves then to a sharper conformation corresponding to a fully bound complex.

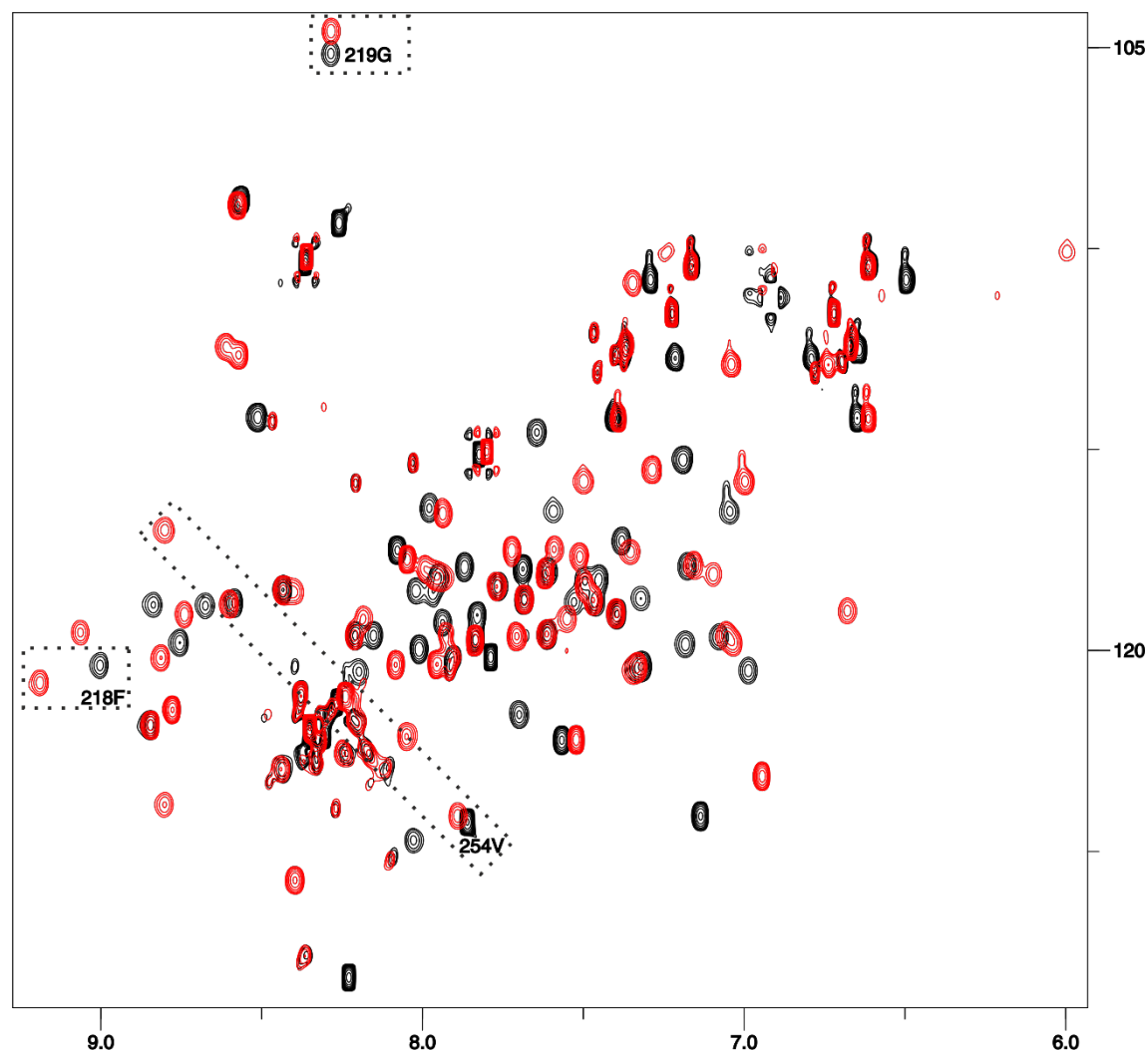


Figure 5.4 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz at 10 °C – in black is presented EB1cΔ8 in the free form (400  $\mu\text{M}$ ) and in red the complex EB1cΔ8 (400  $\mu\text{M}$ ) – 11MACF (600  $\mu\text{M}$ ).

Val254, on the other hand, shows a different profile, where one can observe a very sharp peak in the free form, indicating this residue is in a highly dynamic region in the unbound form. It gradually disappears in this position (decrease of the intensity). The opposite phenomenon is observed for the end position, where the peak gradually appears increasing in intensity until protein saturation, corresponding clearly to a slow exchange regime. In the bound form, this peak is not as sharp as in the free form indicating this region become constrained in terms of flexibility. It is noteworthy that



Val254 at room temperature presented the two states free and bound. At low temperature the same situation is observed with the difference that in the intermediate titration points (0.5 and 0.75  $\mu\text{M}$ ) is now possible to observe free and bound state at the same time.

It is possible that this difference arises from the fact Val254 being located in the C-terminal flexible region of EB1c $\Delta$ 8, whereas Phe218 and Gly219 are closer to the binding site. For the residues located close the binding site the different complexes are observed because they are directly affected by the binding of the ligand. After the ligand binding one assumes the ligand stays in the binding site, and the C-terminus wraps around the ligand, reason why only two sets of peaks can be observed for Val254, there are no intermediate states.

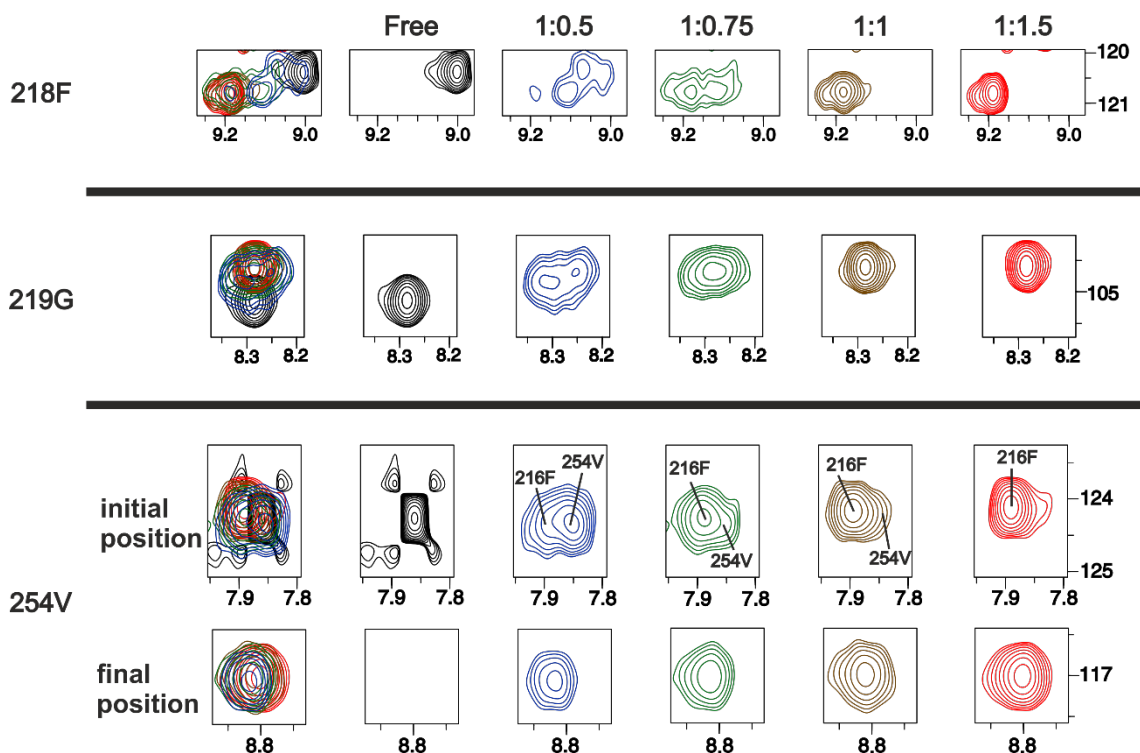


Figure 5.5 - Zoomed regions of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded at 600 MHz, at  $10^\circ\text{C}$  for Phe218 (top panel), Gly219 (middle panel) and Val254 (bottom panel). The bottom panel is divided in two sections, top section shows the region where the initial Val254 peak appears and the bottom section shows the region where the final peak, referring to the bound form appears. First column shows the overlay of all spectra, and the consecutive columns show the spectra referring to EB1c $\Delta$ 8 free at 400  $\mu\text{M}$ , EB1c $\Delta$ 8:11MACF 1:0.5 ratio, 1:0.75 ratio, 1:1 ratio and 1:1.5 ratio, respectively.

Exchange regime shows that binding affinity increases from 4MACF to 11MACF. The use of large concentrations of the shorter length peptides (4MACF and 6MACF respectively) may induce non-specific binding. To address this issue, the next section

will be dedicated to identification of the affected regions upon ligand binding in EB1c's structure.

### 5.1.1 Chemical shift changes and mapping

The mapping of the residues affected upon ligand addition, which resonances present larger chemical shift changes, gives important information on the region(s) of the protein affected upon ligand binding. The chemical shift changes values ( $\Delta\delta$ ) were calculated using Equation 5.1

$$\Delta\delta_{ppm} = \sqrt{((\Delta\delta_{HN})^2 + (\Delta\delta_N * \alpha_N)^2)}, \text{ Equation 5.1}$$

were where,  $\alpha_N$  = Scaling Factor of 0.15.

Upon addition of 4MACF, Tyr247 ( $\Delta\delta = 0.35$  ppm) was the most affected residue, followed by Thr249 ( $\Delta\delta = 0.19$  ppm). Glu213 ( $\Delta\delta = 0.062$  ppm), Phe218 ( $\Delta\delta = 0.065$  ppm), Arg222 ( $\Delta\delta = 0.056$  ppm), Glu225 ( $\Delta\delta = 0.056$  ppm) and Ala248 ( $\Delta\delta = 0.077$  ppm) also shown to be affected by ligand addition to the protein, but to a smaller extent. All these amino acid residues are within the region defined as the binding site for the SxIP motifs, indicating this small peptide fragment has the specificity for EB1c's hydrophobic binding site despite being such a short sequence. Additional CSPs that are not located in the above-mentioned region are observed for Asn235 ( $\Delta\delta = 0.065$  ppm) and Asp236 ( $\Delta\delta = 0.078$  ppm). Despite these two residues being located further away from the binding site but in a dynamic region it is likely that they are not directly affected by the ligand but as an allosteric result. These results prove that despite not being a very strong interaction, the SxIP sequence reveals a very high specificity for the hydrophobic binding site of EBH domain of EB1.

6MACF induces chemical shift changes in a larger number of residues – Figure 5.7. Most of the residues are again located around the binding site for the SxIP motif - Arg214 ( $\Delta\delta = 0.09$  ppm), Phe218 ( $\Delta\delta = 0.11$  ppm), Lys220 ( $\Delta\delta = 0.28$  ppm) and Ala248 ( $\Delta\delta = 0.19$  ppm). The most affected residue is again Tyr247 ( $\Delta\delta = 0.86$  ppm), followed by Thr249 ( $\Delta\delta = 0.48$  ppm) and Val254 ( $\Delta\delta = 0.40$  ppm). Whereas the two first are in agreement with what was observed for 4MACF, Val254 is not. Asp250 ( $\Delta\delta = 0.36$  ppm) and Gly252 ( $\Delta\delta = 0.22$  ppm) give further evidence that the binding region is now more extended than previously. In addition to the extension of the binding

region towards the C-terminus of EB1c, the region immediately below the binding site – Lys212, Glu213 and Arg214, where the additional peptide residues are expected to be located in the complex is affected. This indicates that 6MACF in the complex is located in the similar position as the corresponding part of the full length ligand.

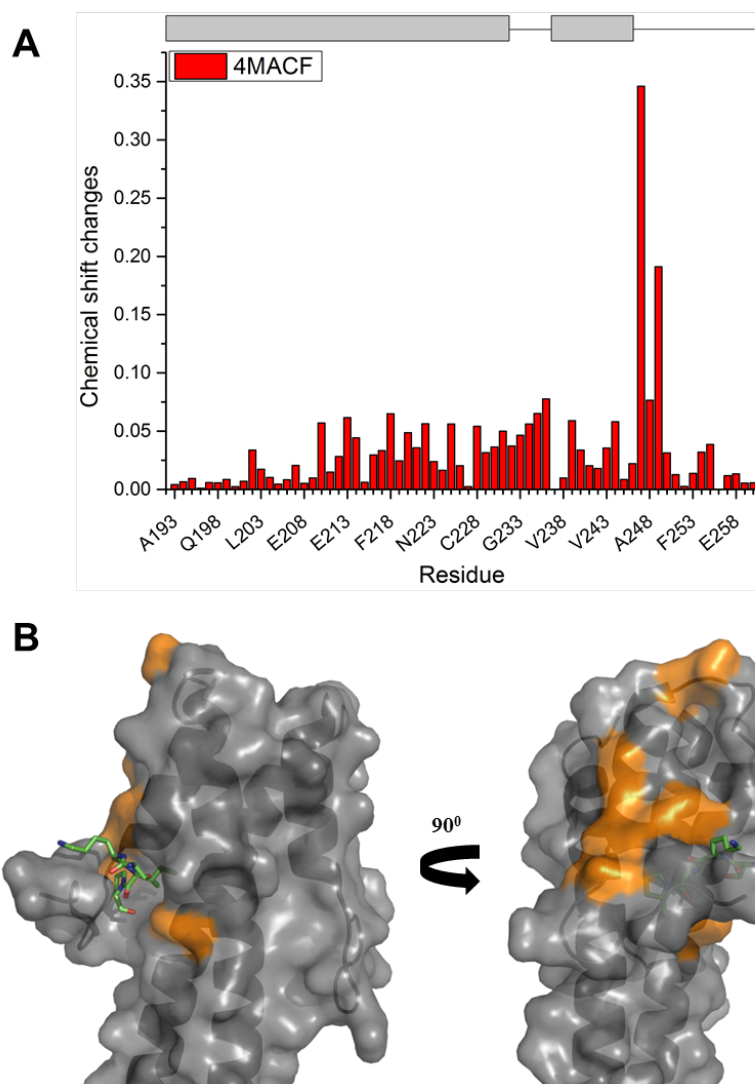


Figure 5.6 – A - Histogram representing the chemical shift perturbations observed between free EB1cΔ8 and bound to 4MACF (SKIP) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

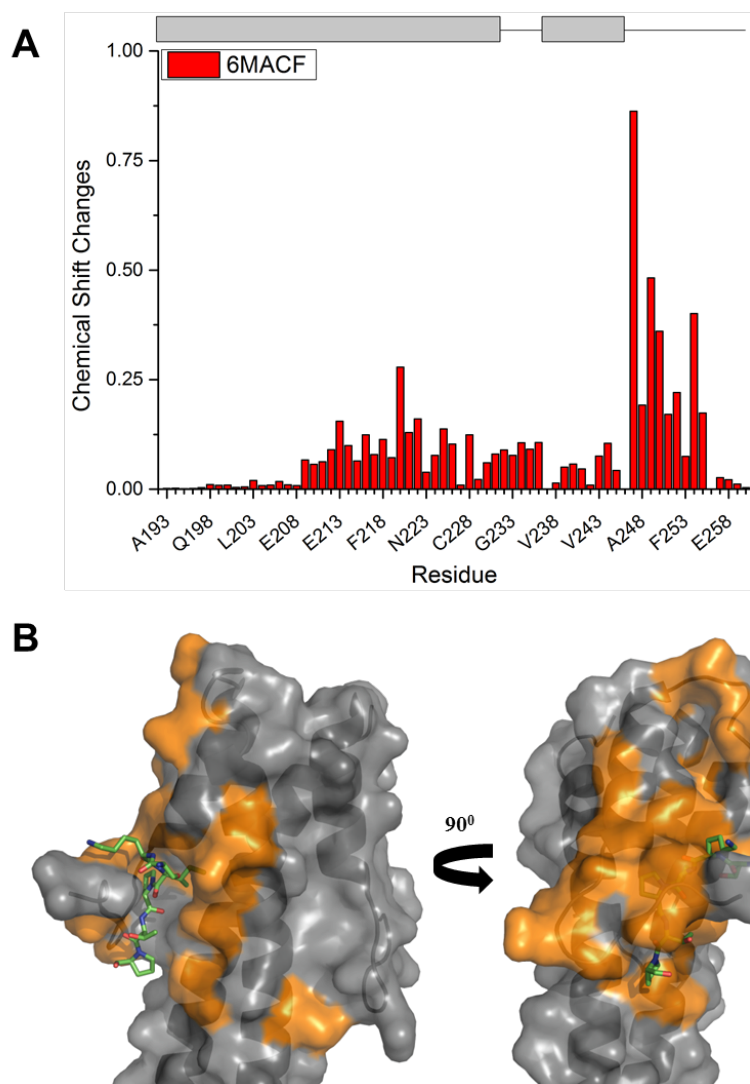


Figure 5.7 – A - Histogram representing the chemical shift perturbations observed between free EB1c $\Delta$ 8 and bound to 6MACF (SKIPTP) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

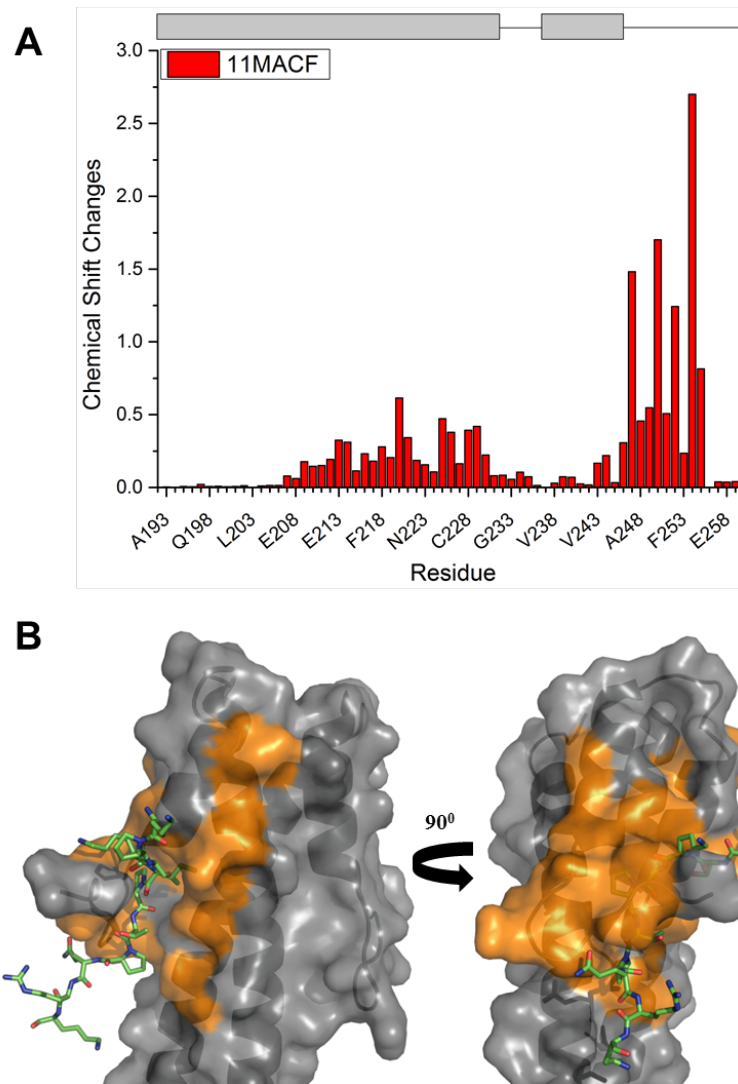


Figure 5.8 – A – Histogram representing the chemical shift perturbations observed between free EB1c $\Delta$ 8 and bound to 11MACF (KPSKIPTPQRK) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

When analysing the chemical shift changes promoted by 11MACF to EB1c $\Delta$ 8 the more immediate difference is, contrarily to what was observed for the shorter sequences, the residue experiencing the largest chemical shift change is not Tyr247 but Val254 ( $\Delta\delta$  of 1.48 and 2.7 ppm respectively), located in the C-terminal tail – Figure 5.8. In this region it is also possible to observe significant chemical shift changes for Asp250, Glu251, Gly252, Phe253 and Ile255. The remaining chemical shift changes are concentrated around the binding site for the SxIP – Phe218, Lys220,

Leu221, Glu225, Leu226, Asp244, Leu246, Tyr247, Ala248, Thr249, immediately below – Glu213 or immediately above – Cys228, Glu230.

In a more general overview of the chemical shift changes plots it is clear that two main regions affected by the interaction with these ligands – the hydrophobic binding site and the C-terminal region, the latter with higher relevance for the stronger ligand – 11MACF. For the first, it is important to mention that the 4MACF sequence (SKIP) is enough to achieve high specificity – no non-specific binding was observed. This is a highly relevant fact that shows that even for protein-protein interactions that usually present large interfaces it is possible to achieve specificity through such a short four residue sequence. Regarding the C-terminal region, the peaks observed for the resonances of this region (250–260) for free EB1cΔ8 are sharper and more intense, indicating more mobility in this region. However, analysis of the intensities of NH cross peaks for the complex formed between EB1cΔ8 and each of the peptides shows that the intensities get attenuated for this region with the increasing length of the peptide – Figure 5.9. This observation correlates with the fact that the stronger binder, 11MACF, promoted chemical shift changes in residues in the C-terminal region, meaning that engaging of the C-terminus region is important to achieve a tighter interaction.

The results obtained point to an increase in the binding affinity from the shorter (4MACF) to the longer peptide (11MACF). The next section will focus on determination of binding affinity values.

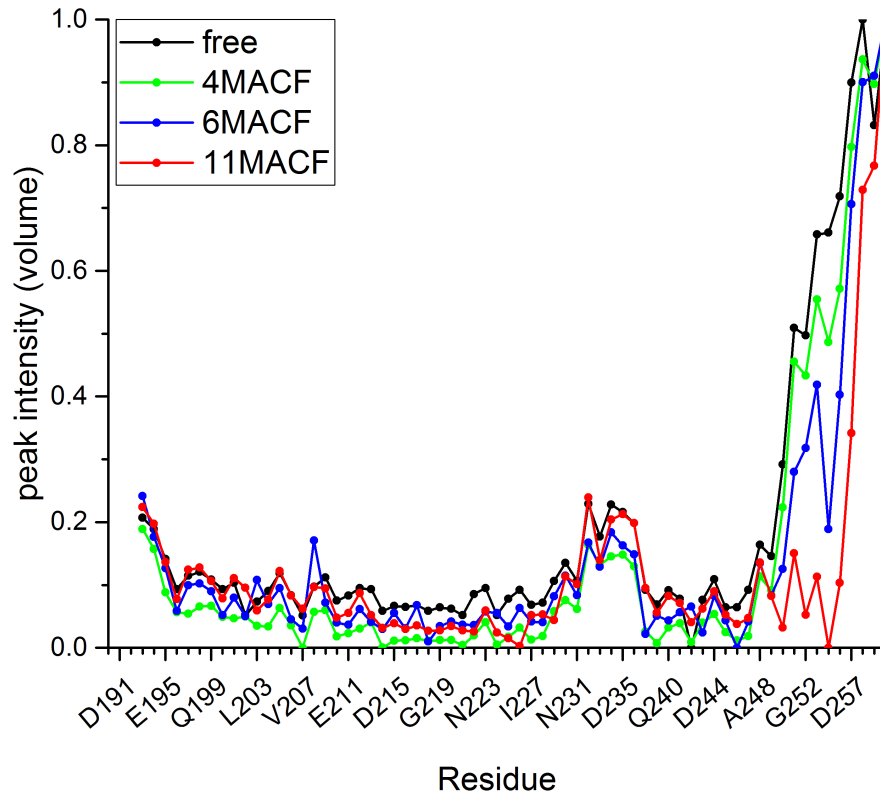


Figure 5.9 – plot of normalised (0-1) NH cross peaks intensities for free EB1cΔ8 (black), final titration point for 4MACF (green), 6MACF (blue) and 11MACF (red).

### 5.1.2 Determination of binding affinity by NMR spectroscopy

The shape of the titration curve (chemical shift *vs* concentration of ligand) can often be fitted straightforwardly to obtain a value for the dissociation constant of the ligand,  $K_d$ . CSP is the only technique that can directly provide both a  $K_d$  value and a binding site from the same set of measurements, more details can be found in Chapter 1 (section 1.2.3.1).<sup>94</sup> Using the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra recorded previously, one can trace and fit the chemical shift changes using the following equation:

$$y = A((B + x) - \sqrt{(B + x)^2 - 4x}), \text{ Equation 5.2}$$

where,  $x = \frac{[L]}{[P]}$ ,  $y = \Delta\delta$ ,  $A = \frac{1}{2}\Delta\delta_{max}$  and  $B = 1 + \left(\frac{K_D}{[P]}\right)$

This equation shows that a good estimate of  $K_d$  can only be obtained if the concentration of protein and ligand is somewhere close to  $K_d$ . The optimum value for the protein concentration is half of the  $K_d$ . Values up to a factor of ten less or more are still usable, as long as the range of concentrations of ligand is large enough, but the error in the fit rises dramatically away from these conditions. In particular, the ligand

should ideally span the range from 0.4 times protein concentration to ten times protein concentration ( $1/5 K_d$  to  $5$  times  $K_d$ ).<sup>94</sup> The lowest concentration of protein that can usefully be employed for HSQC titrations under optimum conditions of sensitivity is currently around  $10 \mu\text{M}$ , implying that  $K_d$  values stronger than about  $1 \mu\text{M}$  are too strong to be fitted from HSQC spectra. At the other end of the scale, CSP data can be used to fit dissociation constants as weak as  $10 \text{ mM}$ , which is close to the upper limit for biologically relevant affinities.<sup>94</sup>

The resulting plot must have significant curvature in order to be able to fit effectively, if there is not enough curvature it means the binding is very far from reaching saturation.

For 4MACF, high concentrations of the ligand were used in order to concentration of the ligand closer to the  $K_d$  value. At the last titration point ( $5 \text{ mM}$ ) the ligand concentration is about half of the  $K_d$ . Although this ligand range of concentrations is too low and therefore not ideal for  $K_d$  determination by NMR, a  $K_d$  value was still estimated from the use of Equation 5.2

As expected, very weak binding is observed and therefore the fitting curve is almost a straight line – Figure 5.10. At the ratios below 1:20 the dependence of chemical changes on ligand concentration is linear, with a small deviation from the straight line at the higher concentrations.

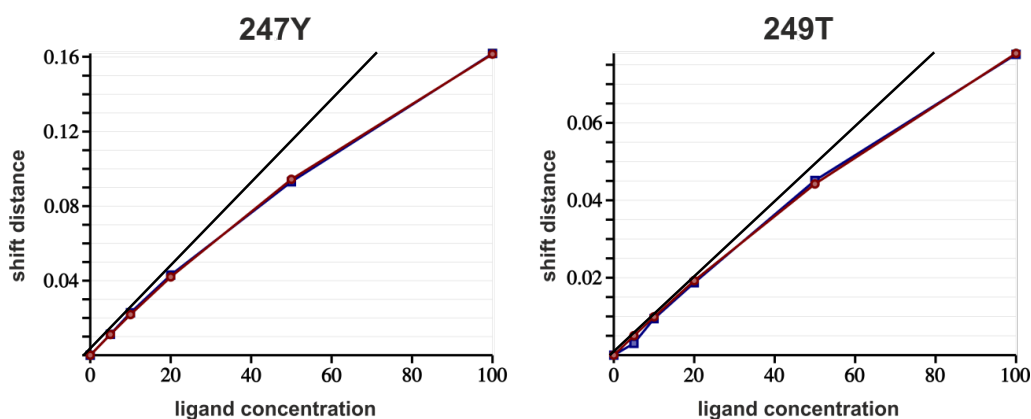


Figure 5.10 – Superposition of the experimental data (blue) and the results of the fitted Equation 5.2 (red) for Tyr247 (left panel) and Thr249 (right panel) of EB1cΔ8 over the course of the 4MACF. These residues show very good agreement between the experimental and fitting function values and hence, were used in the  $K_d$  value determination process. The black straight line was introduced as a reference to facilitate the visualisation of the curvature.



Table 5.1 - The  $K_d$  values calculated for seven residues of EB1c $\Delta$ 8 that best fitted the fitting function during the NMR titration with **4MACF**.

Residue	$K_d$ (mM)
218F	$8.92 \pm 3.81$
220K	$7.74 \pm 2.16$
235N	$9.90 \pm 3.8$
244D	$8.95 \pm 3.23$
247Y	$12.3 \pm 0.91$
248A	$8.7 \pm 2.86$
249T	$16 \pm 3.53$
average	$10.4 \pm 2.90$

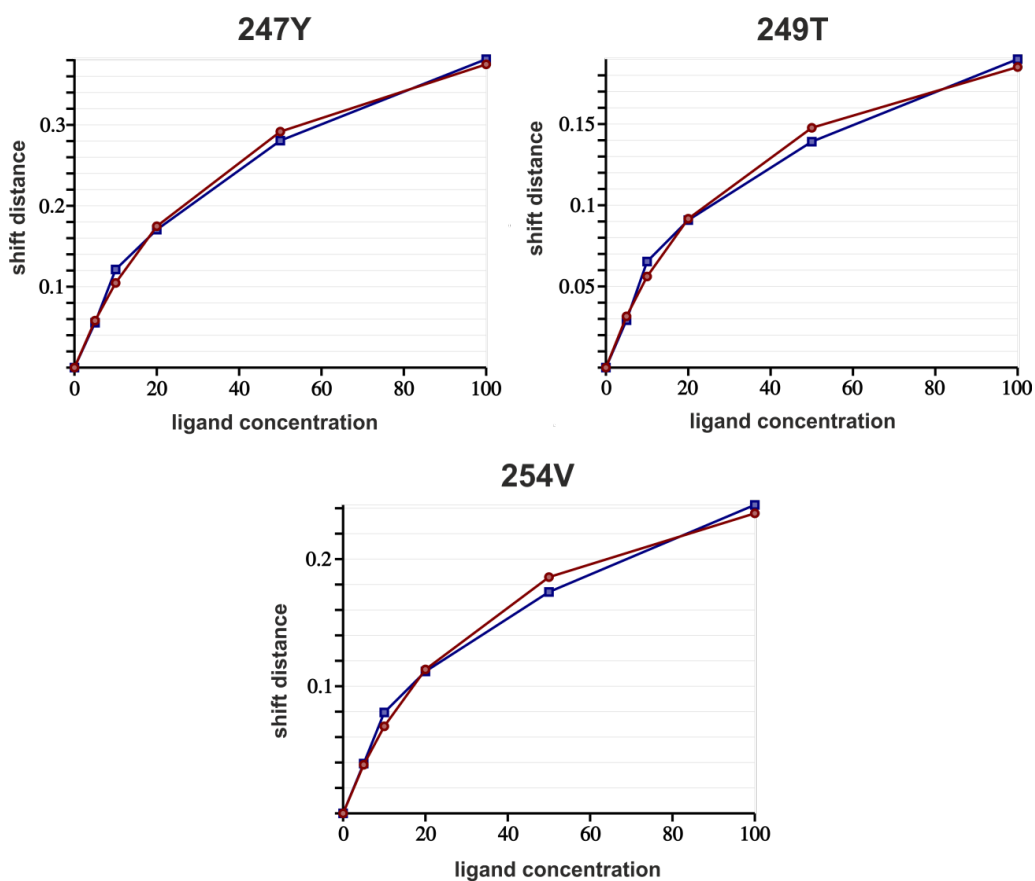


Figure 5.11 - Superposition of the experimental data (blue) and the results of the fitted Equation 1 (red) for Tyr247 (left panel), Thr249 (right panel) and Val254 (bottom centre) of EB1c $\Delta$ 8 over the course of the **6MACF** concentration titration (blue) and the corresponding fitting function (magenta). Similar good agreement between the experimental and fitting function values were observed for all peaks used in the  $K_d$  value determination.

Table 5.2 - The  $K_d$  values calculated for 13 residues of EB1cΔ8 that best fitted the fitting function during the NMR titration with **6MACF**.

Residue	$K_d$ (mM)
212K	$1.69 \pm 0.14$
213E	$2.32 \pm 0.46$
218F	$1.54 \pm 0.11$
221L	$1.45 \pm 0.08$
222R	$1.49 \pm 0.16$
224I	$1.51 \pm 0.14$
247Y	$1.96 \pm 0.26$
248A	$1.67 \pm 0.27$
249T	$1.66 \pm 0.27$
250D	$1.96 \pm 0.39$
252G	$1.58 \pm 0.25$
254V	$1.81 \pm 0.30$
255I	$1.90 \pm 0.33$
average	$1.73 \pm 0.24$

As described previously, this condition is far from optimal for accurate  $K_d$  determination, as the shift dependence does not show enough curvature. However, small systematic deviation from the linear dependence is sufficient for the consistent fitting of the titrations curves for the peaks with the largest chemical shift differences, giving an estimate for the  $K_d$  value of  $\sim 10$  mM (Table 5.1). This value is comparable to the maximum concentration of the peptide used in the titration, further supporting the validity of the approach.

Quantitative analysis of spectral changes for 6MACF titrations (see above) indicated higher affinity of the interaction with EB1cΔ8 than for 4MACF. In agreement with this we detected significant curvature in the titration curves, optimal for the fitting (Figure 5.11). The resulting  $K_d$  values were close for all the peaks with large chemical shift changes that can be reliably followed through the titration (Table 5.2), giving the average value of 1.7 mM for this interaction. This value is significantly lower than the maximum peptide concentration used in the titration, making the titration conditions close to optimal.

It was already shown that 11MACF is the strongest binder of the three tested peptides. Resonances in an intermediate exchange regime are not ideal for  $K_d$  determination using chemical shift changes, due to broadening and disappearance of peaks at certain ratios and complex concentration dependence of chemical shifts. Therefore, the peaks in fast exchange regime with minimal broadening were selected for the  $K_d$  estimation.

The titration curves for these peaks could be fitted well into the fast exchange model (Figure 5.12). However, the variation of the estimated  $K_d$  values was large with large associated errors (Table 5.3). This variation in the  $K_d$  estimates across the peaks indicates that the titration conditions are far from optimal. The average  $K_d$  value of 1.3  $\mu\text{M}$  is much smaller than the protein concentration of 100  $\mu\text{M}$  used in the titration, leading to the limited curvature; additionally, a number of resonances used in the estimation demonstrated different degree of broadening at the intermediate titration points, explaining variations in the shape of the titration curves. All the fitting plots for these residues show clearly that the system reached saturation immediately above 1:1 ratio (Figure 5.12) as expected for the estimated  $K_d$  value, validating the  $K_d$  estimation.

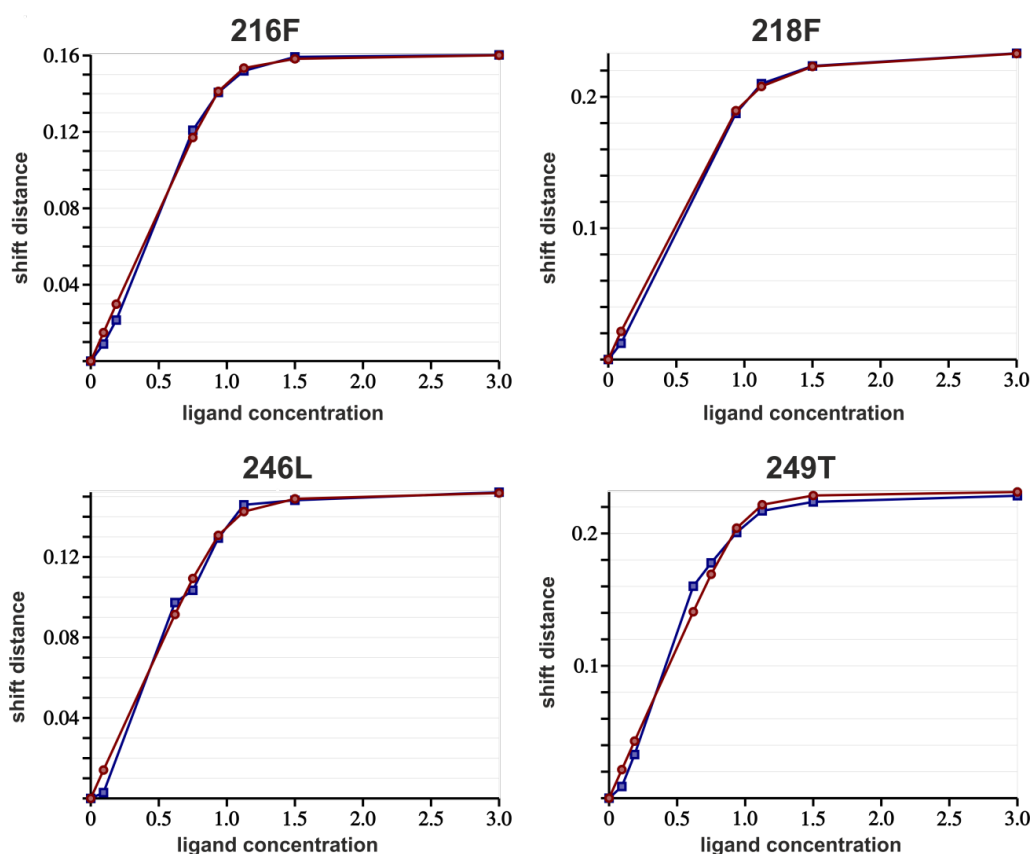


Figure 5.12 - The fit curve produced for Phe216 (left upper panel), Phe218 (right upper panel), Leu246 (left bottom panel) and Thr249 (right bottom panel) of EB1c $\Delta$ 8 over the course of the **11MACF**. These residues show very good agreement between the experimental and fitting function values and hence, were used in the  $K_d$  value determination process. The black straight line was introduced as a reference to facilitate the visualisation of the curvature.

Table 5.3 - The  $K_d$  values calculated for seven residues of EB1c $\Delta$ 8 that best fitted the fitting function

during the NMR titration with **11MACF**.

Residue	$K_d$ ( $\mu\text{M}$ )
216F	$0.822 \pm 0.77$
218F	$3.45 \pm 1.26$
230E	$0.547 \pm 0.92$
244D	$0.717 \pm 1.70$
246L	$1.38 \pm 1.38$
248A	$1.51 \pm 1.86$
249T	$0.818 \pm 1.23$
average	$1.32 \pm 1.30$

6MACF titration is the only one that presents ideal conditions for fitting of Equation 5.2. Despite this it was still possible to estimate  $K_d$  values for 4MACF and 11MACF. The extension of the SxIP motif by two residues, threonine and proline (SKIPTP), improved the binding to EB1c $\Delta$ 8 by roughly 10 fold – 1.7 mM. Further extension, 11MACF (KPSKIPTPQRK) has a remarkable effect in the binding affinity, enhancing it to about 1000 fold – 1.32  $\mu\text{M}$ . Is this enhancement only result of the interaction of the C-terminus as postulated in the beginning of this chapter or are there extra contributions with the coiled coil region? To address this question, the next section will focus on the testing of the same peptides in the absence of the C-terminus.

## **5.2 NMR screening of MACF peptides against EB1c $\Delta$ 16**

From the previous experiments it was possible to observe that a stronger interaction with EB1c $\Delta$ 8 affected the residues around SxIP binding site but also the residues in the C-terminus. To further investigate the role of the C-terminal it was decided to test the same set of ligands – 4MACF, 6MACF and 11MACF, in the absence of the eight C-terminal residues – EB1c $\Delta$ 16. The same approach was used before,  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC titrations, since it involves relatively simple experiments but it gives valuable information on the strength and localisation of the interaction. The removal of the C-terminus had a minor effect on the NMR spectra, with the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum exhibiting highly dispersed cross-peaks of uniform intensity that correspond to a stable coiled coil – section 3.3 – Chapter 3.

For 4MACF - Figure 5.13, the chemical shift changes are overall smaller than the ones observed for the same ligand interaction with EB1c $\Delta$ 8. However, the trend is similar to what was observed for the other construct, with Tyr247 being the most affected

residue, followed by Thr249. Glu225, Phe218 and Arg222, similarly to what was observed before were also affected upon ligand binding. Overall, the chemical shift changes distances were reduced to half of what was observed for EB1c $\Delta$ 8.

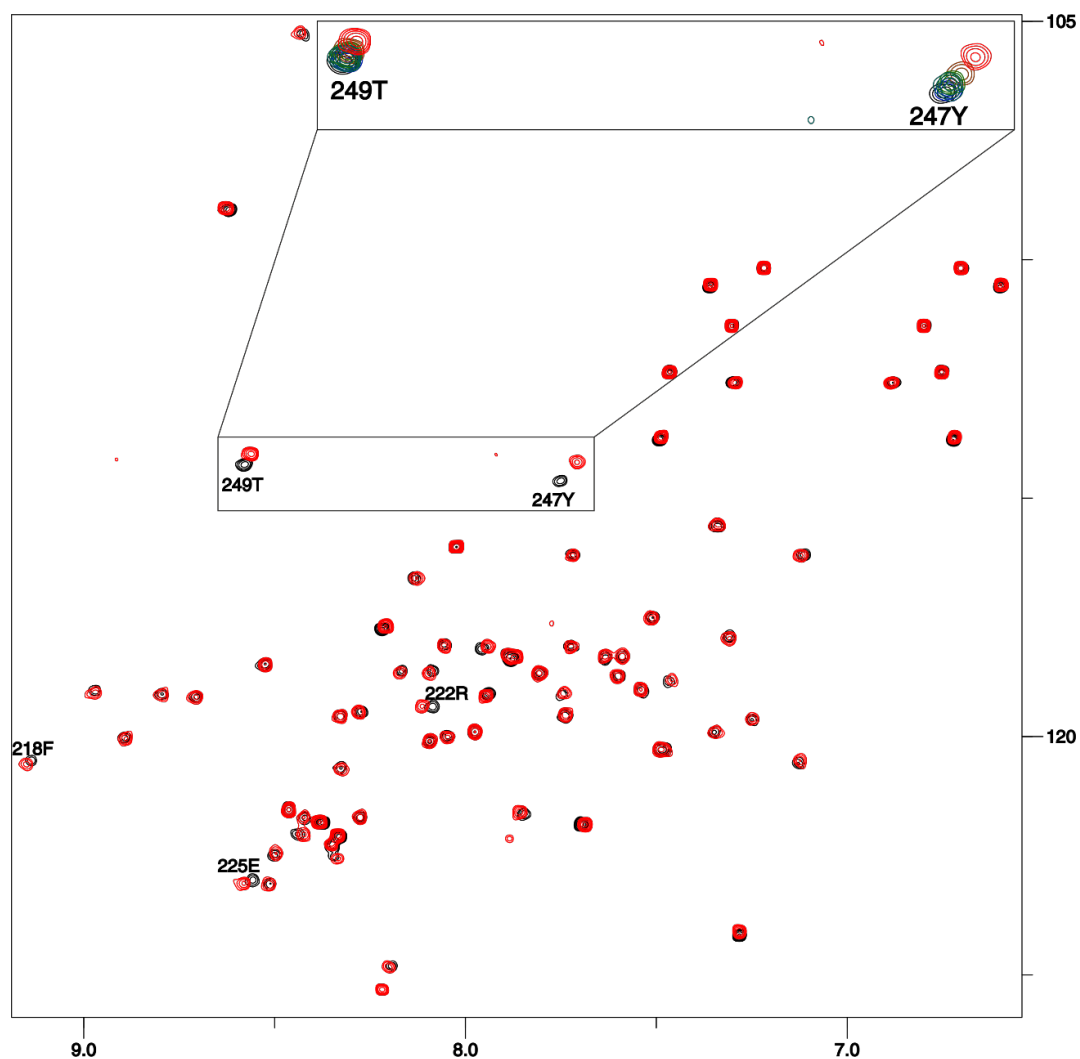


Figure 5.13 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 16 (50  $\mu\text{M}$ ) recorded at 600 MHz - in the free form (black) and in the presence recorded of 4MACF (5000  $\mu\text{M}$ , red) at 600 MHz The insets show regions of interest and the spectra corresponding to the following 4MACF concentrations, 250, 500, 1000, 2500 and 5000  $\mu\text{M}$ .

When EB1c $\Delta$ 16 chemical shift changes are measured upon the addition of 6MACF – Figure 5.14, the chemical shift changes distribution is very similar to the one observed for EB1c $\Delta$ 8, with the exception obviously of Val254, not present in this construct. Therefore, Tyr247 ( $\Delta\delta= 0.45$  ppm) and Thr249 ( $\Delta\delta= 0.29$  ppm) are, once again the most affected residues, followed by Lys220 ( $\Delta\delta= 0.13$  ppm) and Arg222 ( $\Delta\delta= 0.11$  ppm). Glu213 ( $\Delta\delta= 0.09$  ppm), Phe218 ( $\Delta\delta= 0.06$  ppm), Glu225 ( $\Delta\delta= 0.09$  ppm), and Ala248 ( $\Delta\delta= 0.07$  ppm). Similarly to what was observed for the previous ligand the

chemical shift changes distances are about half of the ones observed for EB1cΔ8. Another striking difference is the fact that a subtle peak broadening was observed for resonances such as the ones belonging to Glu213 and Tyr247 when the complex was formed with EB1cΔ8, whereas for this complex all resonances are clearly in a fast exchange regime.

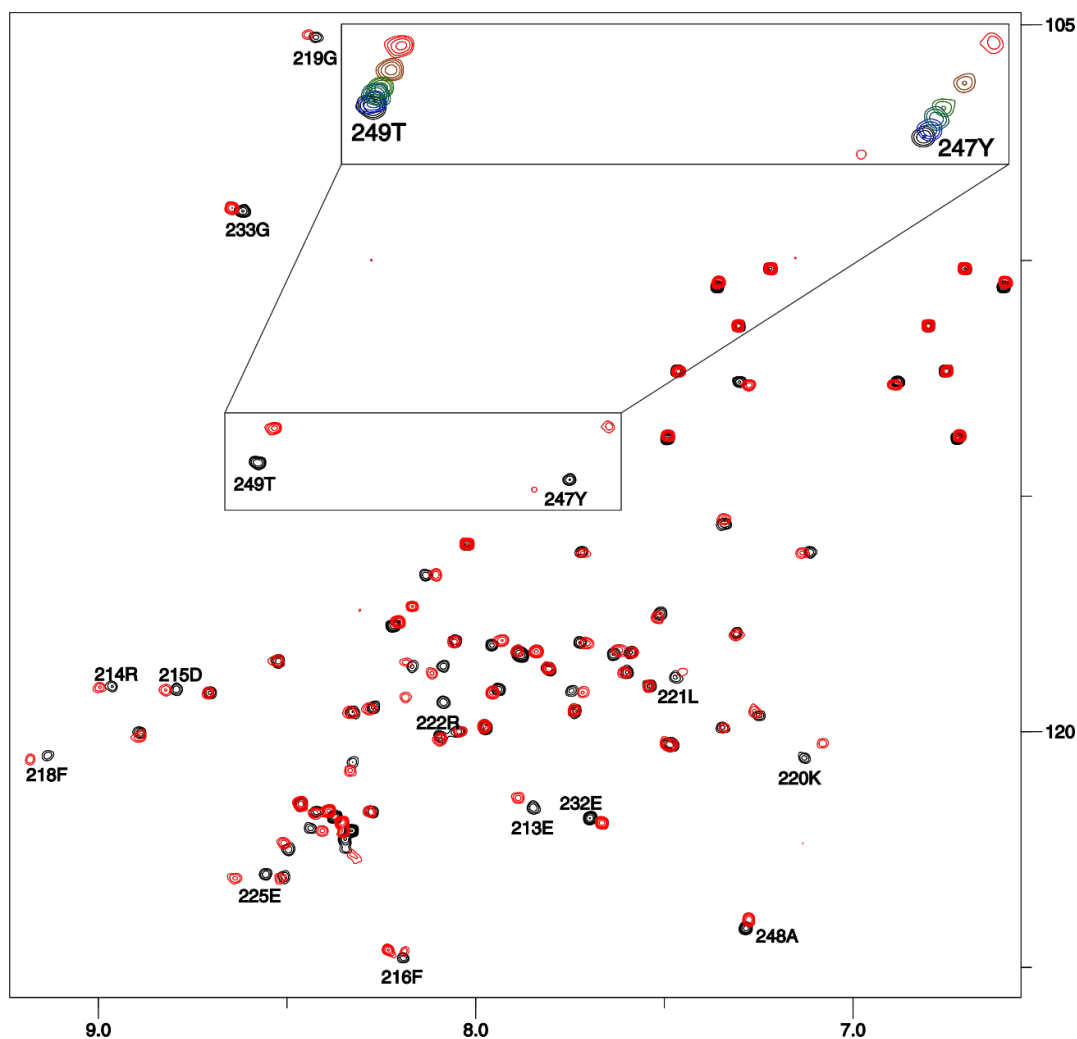


Figure 5.14 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1cΔ16 (50  $\mu\text{M}$ ) recorded at 600 MHz - in the free form (black) and in the presence recorded of 6MACF (5000  $\mu\text{M}$ , red) at 600 MHz. The insets show regions of interest and the spectra corresponding to the following 6MACF concentrations, 250, 500, 1000, 2500 and 5000  $\mu\text{M}$ .

For 11MACF – Figure 5.15, the removal of the C-terminal tail also moves the exchange regime towards fast exchange. This trend is very clear for Ala248 and Thr249. The chemical shift changes are now similar in terms of magnitude when compared with the complex formed with the previous construct. Ala248 shift change is now about five times smaller, Asp250 about two times, Tyr247 about 1.2 times and surprisingly the chemical shift change for Thr249 and Arg222 are now larger than the

ones observed for the complex with EB1cΔ8.

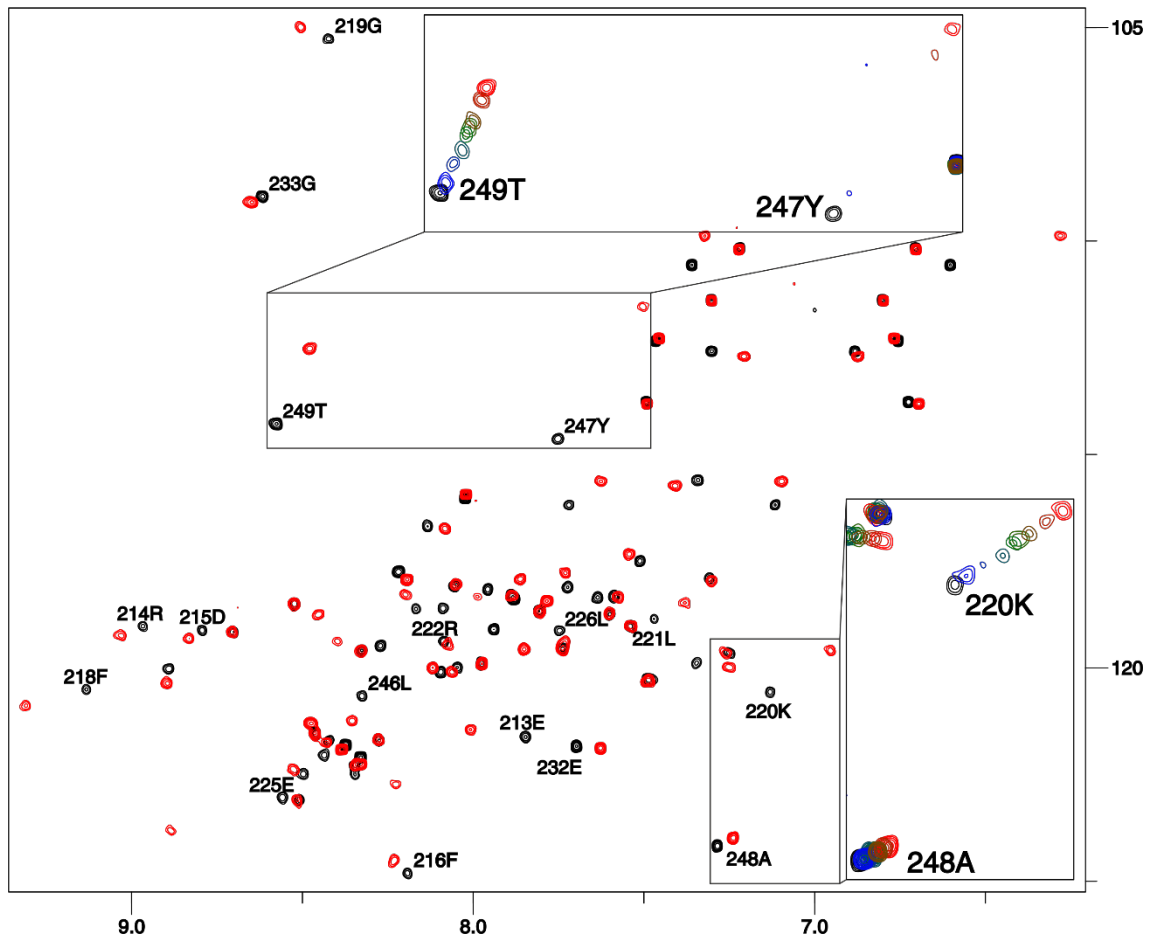


Figure 5.15 - Overlay of  $^1\text{H},^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1cΔ16 (100  $\mu\text{M}$ ) recorded at 600 MHz - in the free form (black) and in the presence recorded of 11MACF (400  $\mu\text{M}$ , red). The insets show regions of interest and the spectra corresponding to the following 11MACF concentrations, 12.5, 25, 75, 82.5, 100, 200 and 400  $\mu\text{M}$ .

The binding is definitely weaker – no significant line broadening and therefore intermediate exchange observed, but the chemical shift changes distances are not as affected as what was observed for the shorter length peptides. The removal of part of the C-terminal flexible tail (residues 253-260) had an obvious effect in the interaction between EB1c and MACF2, diminishing the strength of the interaction as it was expected.

### 5.2.1 Chemical shift changes and mapping

In terms of localisation of the binding all three peptides still target the hydrophobic binding site specific for the interaction with the SxIP motif, and therefore the absence of the C-terminus does not affect the specificity of the SxIP motif for EB1c. On one

hand the C-terminal tail seems to contribute for a stronger interaction with ligands, nevertheless it is the hydrophobic pocket that gives the specificity for the interaction. This was expected since it is widely described in the literature that the specificity for the EB1c – SxIP proteins interaction relies on SxIP motif and its interaction with the hydrophobic pocket. The extra interactions, away from the binding site, only stabilise the interaction.<sup>49</sup> Tyr247 and Thr249 are now the residues more affected by the ligand interaction. The distribution pattern for the chemical shift changes is similar to the one observed before, with the existence of two majorly affected regions – Phe218 – Glu225 and Tyr247-Thr249.

The largest chemical shift change observed for the complex with 4MACF is Tyr247 ( $\Delta\delta = 0.16$  ppm) - Figure 5.16. As described previously the magnitude of the distance is now smaller – previous  $\Delta\delta = 0.35$  pm. Similar trend is observed for Phe218,  $\Delta\delta = 0.03$  vs 0.07 ppm, Arg222,  $\Delta\delta = 0.03$  vs 0.06 ppm, and Glu225,  $\Delta\delta = 0.04$  vs 0.06 ppm.

When bound to 6MACF, EB1c $\Delta$ 16 changes in chemical shift follow the same principle as for 4MACF - Figure 5.17. Tyr247 shows a  $\Delta\delta$  of 0.45 ppm (whereas for EB1c $\Delta$ 8 was of 0.86 ppm). Arg214  $\Delta\delta$  changed from 0.16 to 0.08 ppm, Phe216 0.12 to 0.05 ppm, Lys220 from 0.28 to 0.13 ppm, and finally Ala248 0.19 to 0.07 ppm and Thr249 from 0.48 to 0.28 ppm.

The complex with 11MACF does not show the same pattern as the one observed before, for the simple reason that some of the affected residues are now not present, *e.g.* Val254. Except for that region the chemical shift changes follow the same pattern, and opposite to what was observed for the shorter length peptides not all chemical shift distances were reduced. Thr249 has now a larger chemical shift change –  $\Delta\delta = 0.69$  ppm, whereas previously was 0.55 ppm. Arg222 also has a bigger chemical shift change,  $\Delta\delta = 0.31$  ppm and previously 0.19 ppm. All other resonances seem to have the chemical shift distances reduced but not to such a noticeable magnitude as the one observed for the shorter length peptides. This is a clear effect of the exchange regime shifted from intermediate to fast, the binding is weaker but the chemical shift distances are not so affected.



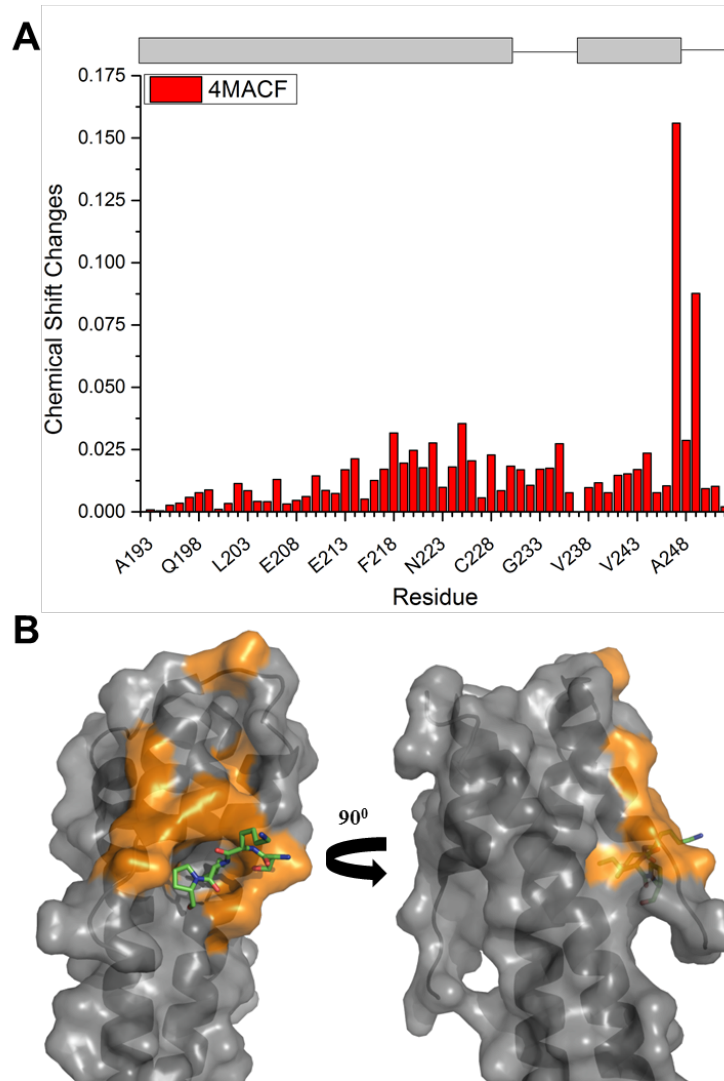


Figure 5.16 - A - Histogram representing the chemical shift perturbations observed between free EB1c $\Delta$ 16 and bound to 4MACF (SKIP) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

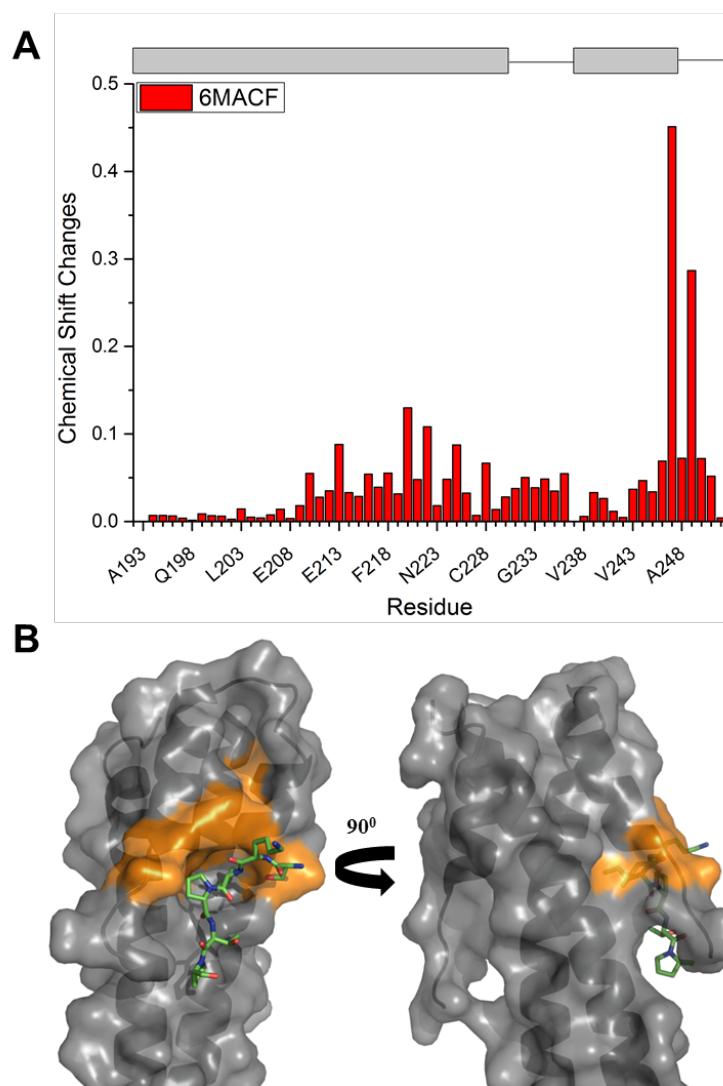


Figure 5.17 - A - Histogram representing the chemical shift perturbations observed between free EB1cΔ16 and bound to 6MACF (SKIPTP) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

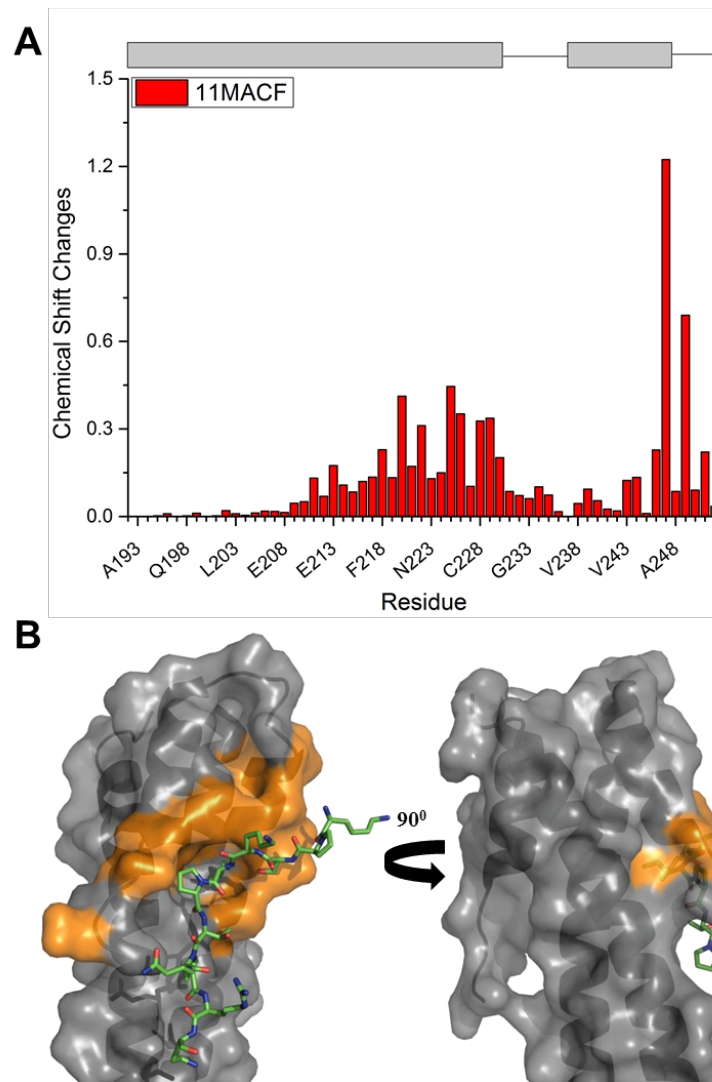


Figure 5.18 – A - Histogram representing the chemical shift perturbations observed between free EB1c $\Delta$ 16 and bound to 11MACF (KPSKIPTPQRK) using the Equation 5.1. The secondary structure of the EB1 is given above the plot. B – Chemical shift changes mapping on the crystal structure (PDB accession code: 3GJO), the selected residues were selected by applying the following formula: remove all residues with chemical shift changes larger than standard deviation ( $\sigma$ ) of the chemical shift changes multiplied by three, re-calculate the standard deviation ( $\sigma$ ) of the chemical shift changes residues and select the residues presenting chemical shift changes above that value.

### 5.2.2 Determination of binding affinity by NMR spectroscopy

Determining the binding affinity for these interactions can be even more challenging than before, as these are even weaker than the ones reported for EB1c $\Delta$ 8. It is noteworthy to reinforce the idea these are only approximations and with this approach one only wants to facilitate the comparison by attributing numerical values to the titrations.

Because for the previous  $K_d$  determination for the 4MACF-EB1c $\Delta$ 8 the value was already on the upper limit for  $K_d$  determination by NMR (10 mM) it was decided to not make an attempt for this interaction between the same ligand and the truncated version of EB1c (EB1c $\Delta$ 16) as the interaction is even weaker and the determined value would not be meaningful.

For 6MACF, it is possible to observe that now the fitting curve presents a less pronounced curvature – Figure 5.19, very similar to the case observed for the interaction 4MACF- EB1c $\Delta$ 8. The estimated  $K_d$  is almost seven times higher than before – Table 5.4, indicating that despite this peptide does not seem to make a large number of interactions with the C-terminus of EB1c, the removal of this region penalises the binding to a large extent. It is possible that the presence of the C-terminus shields the ligand from the solvent, leaving hydrophobic residues such as proline or isoleucine less solvent exposed.

As expected, the same trend is observed for 11MACF- EB1c $\Delta$ 16 – Figure 5.20, Table 5.5. The estimated  $K_d$  is now 12  $\mu$ M, about ten times higher than the one observed previously. Perhaps the lack of the flexible C-terminus is even more noticeable for this interaction as 11MACF CSPs indicated that important changes could be observed for the C-terminal region upon addition of this ligand.

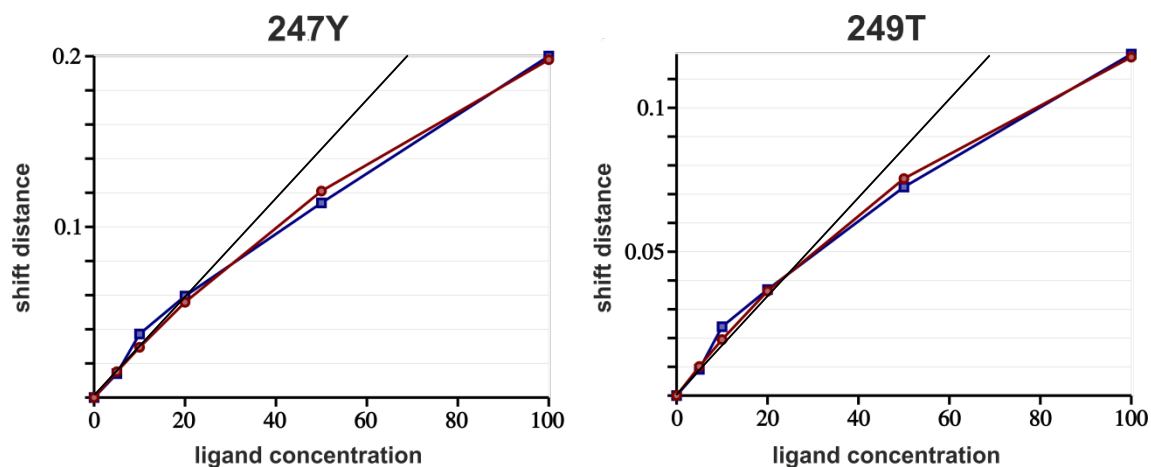


Figure 5.19 - The fit curve produced for 247Y (left panel), 249T (right panel) of EB1c $\Delta$ 16 over the course of the 6MACF concentration titration (blue) and the corresponding fitting function (magenta). These residues show very good agreement between the experimental and fitting function values and hence, were used in the  $K_d$  value determination process. The black straight line was introduced as a reference to facilitate the visualisation of the curvature.

Table 5.4 - The  $K_d$  measured for five residues of EB1c $\Delta$ 16 that best fitted the fitting function during the NMR titration with **6MACF** and hence, were used in the estimation of the binding affinity.

Residue	$K_d$ (mM)
220K	$7.54 \pm 2.24$
222R	$6.97 \pm 2.61$
225E	$5.87 \pm 1.00$
247Y	$8.67 \pm 2.37$
249T	$6.29 \pm 1.17$
average	$7.07 \pm 1.88$

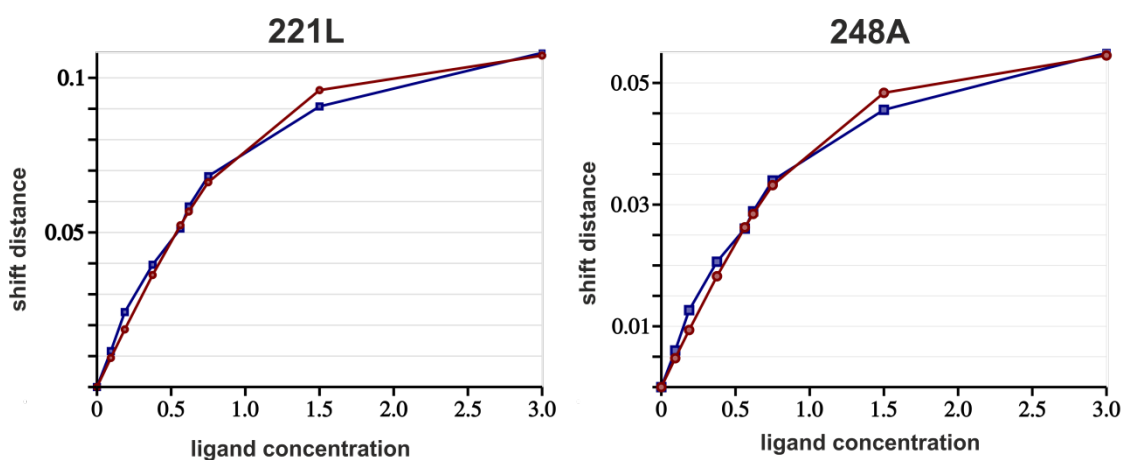


Figure 5.20 - The fit curve produced for 221L (left panel), 248A (right panel) of EB1c $\Delta$ 16 over the course of the **11MACF** concentration titration (blue) and the corresponding fitting function (magenta). These residues show very good agreement between the experimental and fitting function values and hence, were used in the  $K_d$  value determination process.

Table 5.5 - The  $K_d$  measured for four residues of EB1c $\Delta$ 16 that best fitted the fitting function during the NMR titration with **11MACF** and hence, were used in the estimation of the binding affinity.

Residue	$K_d$ ( $\mu$ M)
218F	$9.03 \pm 3.09$
221L	$11.8 \pm 4.87$
246L	$15 \pm 3.74$
248A	$13.3 \pm 5.83$
average	$12.3 \pm 4.38$

### **5.3 Isothermal titration calorimetry binding assays of 11MACF - EB1cΔ8 and 11MACF - EB1cΔ16**

In order to further characterise the complexes formed between 11MACF and EB1cΔ8 and EB1cΔ16, isothermal titration calorimetry was used. The experimental considerations and methods used in this section are described in 7.4.2 – Chapter 7.

As described in section 1.2.3.2 – Chapter I, ITC directly measures the heat released or absorbed during a biomolecular binding event, allowing the simultaneous determination of all thermodynamic parameters, including stoichiometry ( $n$ ), equilibrium association constant ( $K_a$ ),  $\Delta H$  (enthalpy) and  $\Delta S$  (entropy) for that particular binding interaction. These thermodynamic parameters can be used to calculate the Gibbs free energy ( $\Delta G$ ) for a particular interaction between two macromolecules (Equation 5.3 and 5.4), where a negative value indicates that non-covalent association is occurring. The binding affinity ( $K_a$ ), is inverse to the dissociation constant ( $K_d$ ), therefore smaller  $K_d$  tighter the binding (Equation 5.5). Obtaining all these parameters from one single method makes ITC an extremely useful biophysical technique for determination and comparison of biomolecular interactions.<sup>169,170</sup>

$$\Delta G = \Delta H - T\Delta S, \text{ Equation 5.3}$$

$$K_a = e^{-\Delta G/RT} \leftrightarrow \Delta G = -RT \ln K_a, \text{ Equation 5.4}$$

$$K_d = \frac{1}{K_a}, \text{ Equation 5.5}$$

The interaction with 4MACF and 6MACF was too weak to be measured by ITC. However, the same is not true for the interactions of EB1c and the 11mer peptide – 11MACF. At this stage it would be important to calculate accurately the  $K_d$  values for both EB1cΔ8 and EB1cΔ16 interactions with 11MACF. In addition, and making use of the versatility of this method to calculate parameters such as  $\Delta H$  and  $\Delta S$ , it would be of interest to know how the presence of the C-terminal flexible tail affects these parameters. The individual isotherms for these interactions are shown in Figure 5.21 and Figure 5.22. A total of three experiments were used to calculate the thermodynamic parameters and respective associated errors showed in Table 5.6.

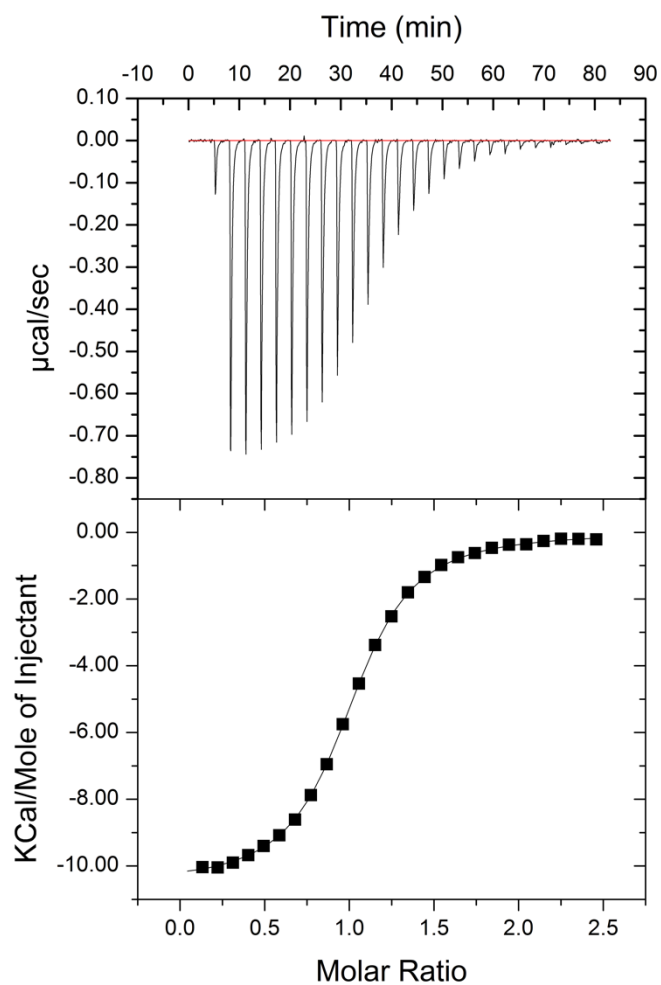


Figure 5.21 - The ITC isotherm (top) and resultant curve (bottom) produced by the ITC binding experiment between  $50 \mu\text{M}$  EB1c $\Delta$ 8 (cell) and  $750 \mu\text{M}$  11MACF (syringe) in 20 mM phosphate, pH 6.5, 50 mM NaCl,  $0.5 \mu\text{M}$  TCEP, 0.02%  $\text{NaN}_3$  on an iTC<sub>200</sub> Microcalorimeter (MicroCal) at 25°C. Fitting of the curve produced by the ITC experiment to a single set of sites curve-fitting model, using Origin7, resulted in a binding affinity ( $K_d$ ) of  $2 \mu\text{M}$ .

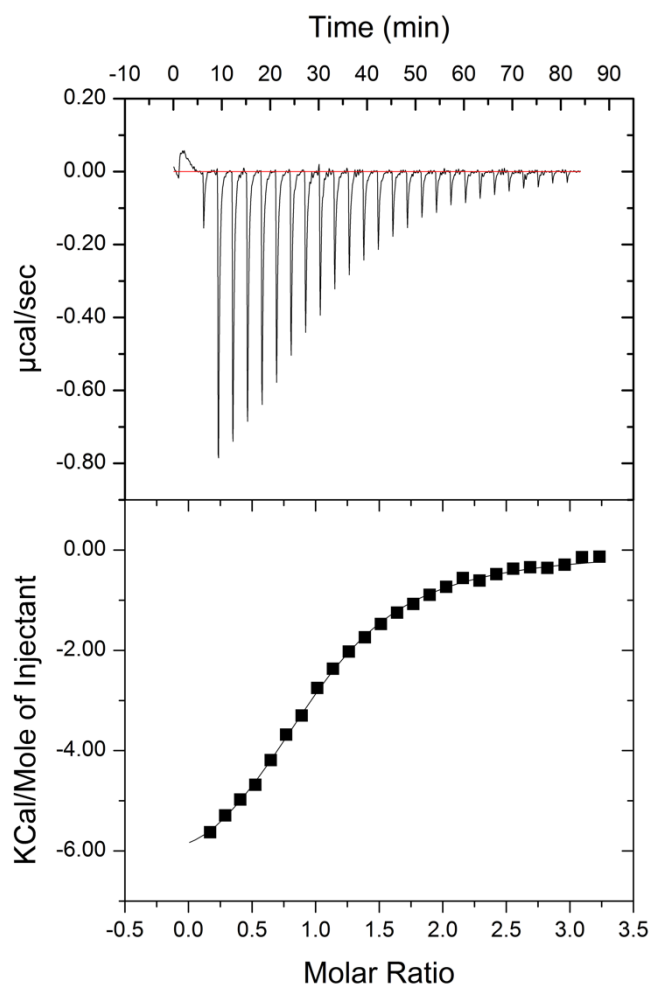


Figure 5.22 - The ITC isotherm (top) and resultant curve (bottom) produced by the ITC binding experiment between 150  $\mu\text{M}$  EB1c $\Delta$ 16 (cell) and 2250  $\mu\text{M}$  11MACF (syringe) in 20 mM phosphate, pH 6.5, 50 mM NaCl, 0.5  $\mu\text{M}$  TCEP, 0.02%  $\text{NaN}_3$  on an iTC<sub>200</sub> Microcalorimeter (MicroCal) at 25°C. Fitting of the curve produced by the ITC experiment to a single set of sites curve-fitting model, using Origin7, resulted in a binding affinity ( $K_d$ ) of 26  $\mu\text{M}$ .

Table 5.6 - Thermodynamic parameters obtained for three ITC binding experiments to detect the interactions of 11MACF with EB1c $\Delta$ 8 and EB1c $\Delta$ 16. Values shown in this table correspond to the average and associated error obtained for the three experiments.

Interaction	$K_d$ ( $\mu\text{M}$ )	$\Delta\text{H}$ ( $\text{kcal mol}^{-1}$ )	$-\text{T}\Delta\text{S}$ ( $\text{kcal mol}^{-1}$ )	$\Delta\text{G}$ ( $\text{kcal mol}^{-1}$ )
EB1c $\Delta$ 8 11MACF	$3.80 \pm 0.82$	$-10.57 \pm 0.63$	$3.15 \pm 0.64$	$-7.42 \pm 0.12$
EB1c $\Delta$ 16 11MACF	$41.5 \pm 8.84$	$-7.11 \pm 0.44$	$1.11 \pm 0.45$	$-6.00 \pm 0.12$



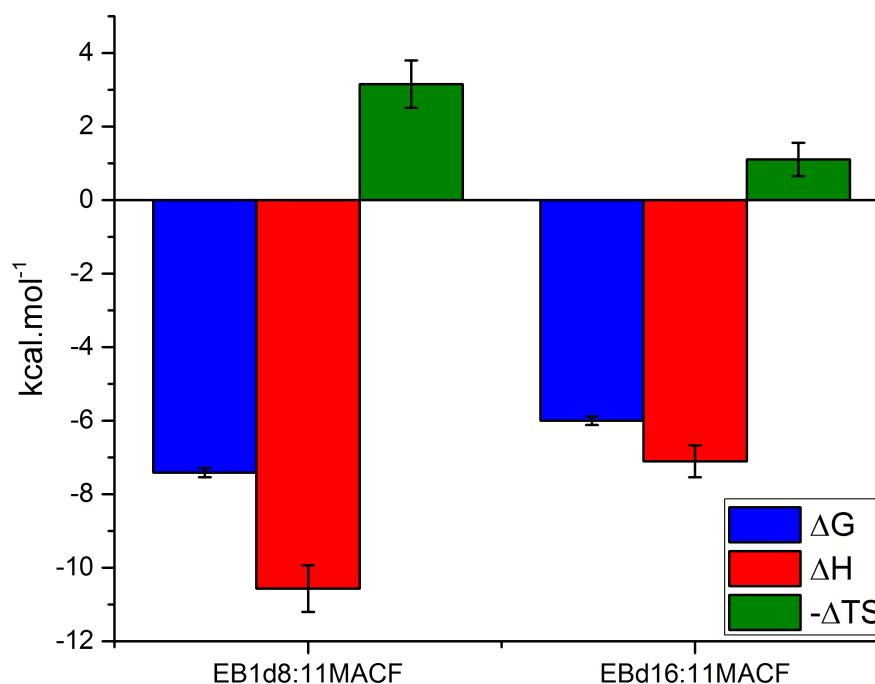


Figure 5.23 - The thermodynamic parameters ( $\Delta G$  (blue),  $\Delta H$  (red) and  $-\Delta TS$  (green)) obtained from the binding interaction between the 11MACF and EB1c $\Delta 8$  (EB1d8, left hand side) and EB1c $\Delta 16$  (EB1d16, right hand side).

The interaction between EB1c $\Delta 8$  and the 11-residue peptide derived from MACF2, 11MACF, it is in the low  $\mu\text{M}$  range,  $3.80 \pm 0.82 \mu\text{M}$ . This value is in agreement to what was reported previously for a longer peptide sequence ( $\sim 26$  amino acid residues) including this sequence,  $3.5 \mu\text{M}$ . This value was obtained using fluorescence polarisation (FP) displacement.<sup>17</sup> As expected, and in accordance with what was observed by NMR, the  $K_d$  for the interaction of the same peptide and EB1c $\Delta 16$  (lacking eight C-terminal residues) is higher, indicating a weaker interaction –  $41.5 \pm 8.84 \mu\text{M}$ , about 10 fold weaker. Once again, this is in agreement with the results obtained by NMR, where the estimated  $K_d$  changed from  $1.32 \mu\text{M}$  to  $12.3 \mu\text{M}$ .

A similar effect was demonstrated by Honnappa and co-workers<sup>31</sup> where a different SxIP containing peptide derived from APC was tested for interaction with EB1c lacking the 20 C-terminal residues (191-247). Because these authors removed important residues part of the binding pocket such as Ala248 and Thr249, the effect is more extreme with no binding observed.

In terms of thermodynamic parameters contribution, Figure 5.23 represents a graphical output of the values present in Table 5.6.  $\Delta G$  is made up of two different contributions,  $\Delta H$  and  $\Delta S$ , and different combination of these values, in principle, elicit the same

binding affinity (*i.e.* the same  $\Delta G$  and therefore the same  $K_a$ ). The binding enthalpy ( $\Delta H$ ) primarily reflects the strength of the interactions of the ligand with the target protein (*e.g.* van der Waals, hydrogen bonds, etc.) relative to those existing with the solvent. The entropy change ( $\Delta S$ ), on the other hand, mainly reflects two contributions: changes in solvation entropy and changes in conformational entropy. Upon binding, desolvation at the binding interface occurs, water is released to bulk solution and a gain in solvent entropy is observed. This gain is particularly important for hydrophobic groups. At the same time, the ligand and certain groups in the protein lose conformational freedom, resulting in a negative change in conformational entropy.<sup>17</sup>> The removal of eight residues from EB1c's C-terminal tail caused enthalpic contributions to be smaller, meaning that it is possible that hydrogen bonds and other non-covalent interactions formed between this region and 11MACF may have been lost. On the other hand, entropic contributions are more favourable when the majority of the C-terminal tail is absent. This may be related with when the C-terminal region is present it tends to become more structured upon binding to SxIP containing peptide. This observation was already described in this thesis – Figure 5.9. This loss of flexibility may cause unfavourable contributions to the system entropy. Despite this, the enthalpic loss for this complex outweighs the entropic gains due to the removal of the flexible region.

The X-ray crystal structure published by Honnappa and co-workers in 2009<sup>31</sup> is the only structure published to date where part of the flexible C-terminus is observed, meaning this region should in principle be less dynamic when EB1c is bound to a peptide derived from a natural ligand. In order to get extra information on how EB1c binds to a SxIP containing peptide, it was decided to elucidate the structure in solution of the complex formed between EB1c $\Delta 8$  and 11MACF. This structure differs from the one supra in two main aspects, the first is a structural model in solution, the second is the fact the SxIP containing peptide is a shorter sequence, 11 residues instead of 30.

#### **5.4 Structural elucidation of the complex EB1c $\Delta 8$ -11MACF using solution NMR**

In this thesis two solution NMR structures of EB1c $\Delta 8$  were already described for the free and bound form – Chapter 3. The structure described in this section aims to

represent the structural conformations of EB1cΔ8 when bound to a longer and higher affinity ligand. The first difference between this complex and the one previously elucidated between EB1cΔ8 and molecule **1a** is the fact this complex is in intermediate exchange and therefore it was not possible to follow the chemical shift changes during the course of the titration from the resonances obtained for free EB1cΔ8. Therefore, all spectra were assigned in an identical manner to the one described for EB1cΔ8 in the free form – Chapter 3. The fact that the complex presents characteristics of an intermediate exchange regime, with extensive line broadening for some of the protein resonances, it was necessary to add excess of ligand (10:11 ratio) in order to get sharper peaks for the ligand without loss (broadening) of the protein resonances – Appendix (A.3).

One hundred and eighty five intermolecular distance restraints were calculated using CcpNmr Analysis<sup>17</sup> and introduced into the calculations of the complex. One hundred and ten distance restraints were iteratively assigned by ARIA<sup>174</sup> yielding a total of 298 distance restraints between the protein and the peptide. Additionally, 209 intramolecular distance restraints for the peptide were also assigned and introduced – Table 5.7. This approach yielded a structure of a complex shown in Figure 5.24. Overall EB1cΔ8 maintains the coiled-coil structure composed by a leucine zipper and four helical bundle in the upper region. The most remarkable difference between this structure and the two previously reported (Chapter 3) is that the C-terminal tail is now in a stable and fixed conformation.

Table 5.7 - NMR restraints and structure statistics for the structures of EB1cΔ8 in complex with 11MACF.

EB1cΔ8-11MACF	
Total restraints used	
NOE restraints*	
<i>All</i>	3863
<i>Protein-ligand</i>	298
<i>Intermonomer</i>	924
<i>Intrapeptide</i>	209
<i>Intraresidue</i>	1133
<i>Sequential</i> ( $ i-j  = 1$ )	968
<i>Medium</i> ( $1 <  i-j  \leq 4$ )	1311
<i>Long range</i> ( $ i-j  > 4$ )	177
Dihedral	
$\phi$ angles	65
$\psi$ angles	65
Hydrogen bonds	
	90
Structure statistics	
Violations	
<i>Distance</i> ( $> 0.5 \text{ \AA}$ )	42 ( $\pm 5$ )
<i>Dihedral angle</i> ( $> 5^\circ$ )	7 ( $\pm 1$ )
Energies (cal/mol)	
<i>Overall</i>	-2179 ( $\pm 208$ )
<i>Bond</i>	119 ( $\pm 9$ )
<i>Angle</i>	479 ( $\pm 21$ )
<i>Improper</i>	242 ( $\pm 29$ )
<i>Dihedral</i>	868 ( $\pm 14$ )
<i>Van der Waals</i>	-48 ( $\pm 28$ )
<i>Electrostatic</i>	-5903 ( $\pm 98$ )
<i>NOE</i>	1906 ( $\pm 119$ )
Geometry – average Values	
<i>Bond</i>	$7.40 \times 10^{-3}$ ( $\pm 5.7 \times 10^{-4}$ )
<i>Angle</i>	$0.91$ ( $\pm 9.96 \times 10^{-2}$ )
<i>Improper</i>	2.50 ( $\pm 0.30$ )
<i>Dihedral</i>	41.56 ( $\pm 0.24$ )
<i>Van der Waals</i>	428.93 ( $\pm 83.98$ )
Average pairwise RMSD ( $\text{\AA}$ )**	
<i>Heavy atoms</i>	2.41 ( $\pm 1.25$ )
<i>Heavy atoms – helical region</i>	1.51 ( $\pm 0.99$ )
<i>Backbone</i>	2.05 ( $\pm 1.33$ )
<i>Backbone – helical region</i>	1.08 ( $\pm 0.93$ )
Ramachandran statistics (%)***	
<i>Most favoured regions</i>	87.0 (99.5)
<i>Additional allowed regions</i>	10.8 (0.3)
<i>Generously allowed regions</i>	1.0 (0.2)
<i>Disallowed regions</i>	1.1 (0)

\*Number in brackets corresponds to the restraints assigned manually

\*\*Helical region corresponds to residues: Glu192-Glu230 and Pro237-Tyr247

\*\*\* Values within brackets correspond to residues Glu192-Glu230 and Pro237-Tyr247 (helical region)

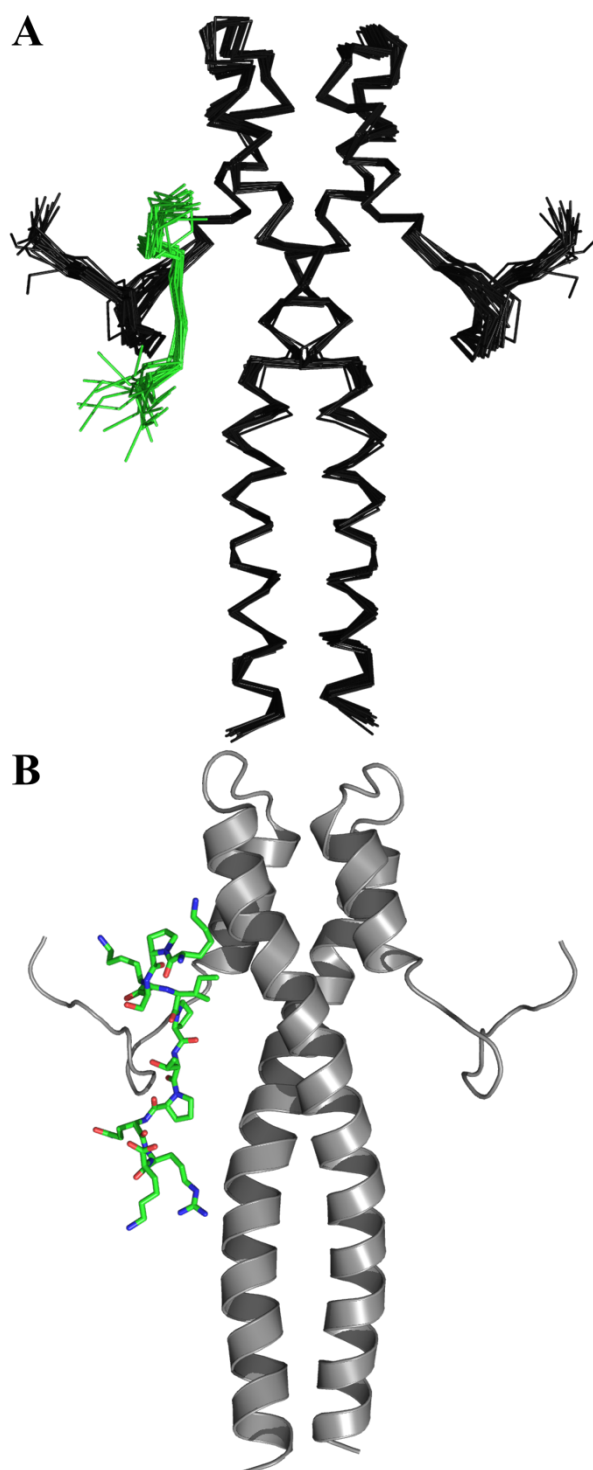


Figure 5.24 - A – Superimposition of the ensemble of 20 best structures of EB1cΔ8 domain in complex with 11MACF. B – Cartoon representation of the lowest energy structure for EB1cΔ8 in complex with 11MACF shown in sticks, where green represents carbon, blue nitrogen and red oxygen.

The solution NMR structure of EB1cΔ8 in the unbound state elucidated the side chain mobility of two residues in the binding site, Arg222 and Tyr247. In the bound form to molecule **1a**, these residues remained in a stable and fixed conformation. This was further proved by the existence of NOE cross peaks between the molecule and both residues of the protein. Similarly, for the complex formed between EB1cΔ8 and 11MACF these residues remain in a stable conformation, with only little side chain variability for Arg222. Again, this is supported by the existence of NOE distance restraints - Table 5.8.

Table 5.8 – NOE calculated distance restraints obtained for Arg222

NOE distance restraints assigned and introduced in the calculations via CcpNmr Analysis						
A	Arg222	H $\alpha$	C	Ile5479	H $\delta$ 1	4.3 Å
A	Arg222	H $\alpha$	C	Ile5479	H $\delta$ 1	4.6 Å
A	Arg222	H $\delta$ 1	C	Ile5479	H $\delta$ 1	4.9 Å
NOE distance restraints calculated by ARIA						
A	Arg222	H $\alpha$	C	Ile5479	H $\delta$ 1	2.9 Å
A	Arg222	H $\delta$ 1	C	Ile5479	H $\delta$ 1	4.2 Å
A	Arg222	H $\beta$ 1	C	Ile5479	H $\delta$ 1	4.5 Å

Table 5.9 - Parameters calculated via CASTp<sup>15</sup> webserver for the hydrophobic pocket shape and size analysis.

structure	Number of mouth openings	Area (Å <sup>2</sup> )		Volume (Å <sup>3</sup> )		sum arc length	corner points
		Solvent accessible	Molecular surface	Solvent accessible	Molecular surface		
1	1	328.791	434.47	320.387	855.13	264.74	110
2	1	184.29	273.09	173.877	493.53	152.3	74
3	2	382.843	519.29	445.541	1067.72	319.25	139
4	2	244.428	351.41	198.826	612.02	200.48	105
5	1	292.415	402.28	290.323	779.39	258.6	118
6	1	239.808	353.61	209.158	623.31	211.9	106
7	1	213.99	359.06	185.383	571.63	184.89	101
8	2	245.327	390.99	195.283	627.09	220.03	111
9	1	200.696	311.21	183.17	531.71	169.54	89
10	1	237.233	366.56	209.798	617.6	185.16	95
11	1	320.497	494.99	244.091	802.65	285.4	146
12	1	315.463	426.35	287.722	805.96	255.98	124
13	1	232.901	351.45	214.77	621.19	205.11	100
14	1	187.433	265.69	200.472	515.31	147.34	70
15	1	226.224	360.73	220.101	619.27	191.32	102
16	1	182.231	266.67	159.245	468.69	146.49	73
17	1	369.165	532.34	389.309	991.42	290.39	141
18	1	301.724	397.48	326.503	817.38	245.4	114
19	1	303.93	424.31	320.598	824.42	250.43	126
20	1	285.503	384.62	325.76	782.8	216.56	107
Average	1.2	264.7446	383.33	255.016	701.411	220.07	108

Standard deviation	0.4	58.87961	73.83241	76.783	161.532	48.581	20.9
--------------------	-----	----------	----------	--------	---------	--------	------

There are three structures where the arginine side chain occupies a more interior position (1, 2 and 4) and these still present large binding pocket volumes (Table 5.9), opposing to what was observed in the free form.

Remarkably, the average binding pocket size for this complex is about five times larger than the average obtained for the complex with **1a** –  $49 \text{ \AA}^3 (\pm 6 \text{ \AA}^3)$  versus  $255 \text{ \AA}^3 (\pm 77 \text{ \AA}^3)$ . The main reason this major difference in the binding site is the contribution of the C-terminal tail of EB1c $\Delta$ 8 (residues 248-260). The C-terminus folds around the peptide protecting it from the solvent and therefore extending the binding site - Figure 5.25.

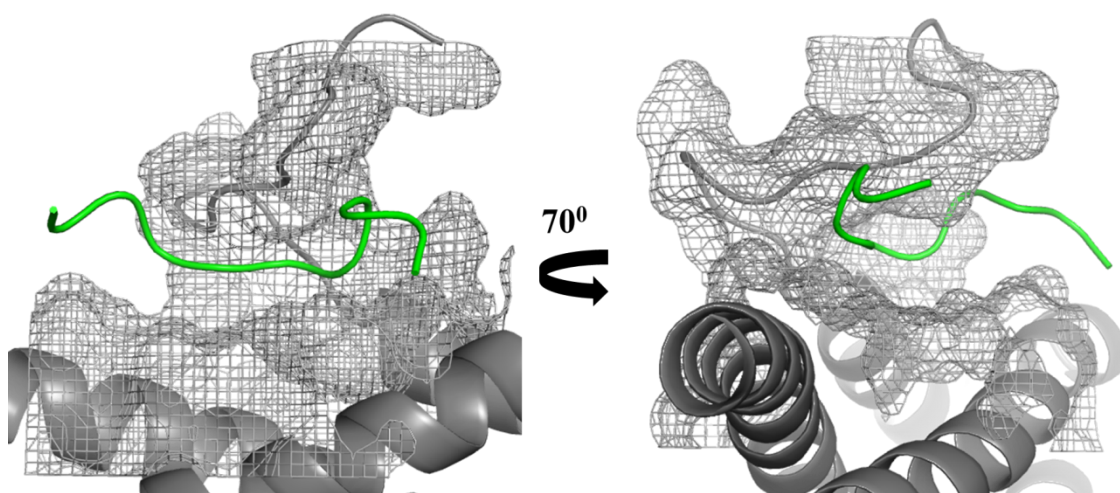


Figure 5.25 – Representation of the binding site of the lowest energy structure (structure 1) obtained for the complex EB1c $\Delta$ 8-11MACF. Protein and ligand are both in cartoon representation, grey and green respectively. The binding pocket is represented as a grey mesh and was calculated using CASTp webserver (solvent accessible volume of  $320 \text{ \AA}^3$ ).

This conformation is observed across the ensemble of 20 calculated structures for this complex and further proved by the existence of intermolecular NOEs between the ligand and the C-terminus of EB1c $\Delta$ 8. These contacts include Val254 to the sequence  $^{5480}\text{PTPQR}^{5484}$ , Ile255 to the sequence  $^{5478}\text{KI}^{\text{TP}}^{\text{PQR}}$ , Pro256 to the sequence  $^{5478}\text{KI}^{\text{TP}}^{\text{PQR}}$ , and finally Asp257 to Lys5478. Finally, Phe253 establishes most of the contacts with Pro5482, with fewer for Ile5479 and Thr5481. A detailed list of the distance restraints derived from the NOE data can be found in Appendix A.3. As mentioned above the C-terminus of EB1c $\Delta$ 8 interacts with the peptide – residues  $^{5478}\text{KI}^{\text{TP}}^{\text{PQR}}$ , but the majority of the contacts are within the region  $^{5478}\text{KI}^{\text{TP}}^{\text{PQR}}$ . Peptide side chains are therefore more restrained for this region and show good

superimposition - Figure 5.26.

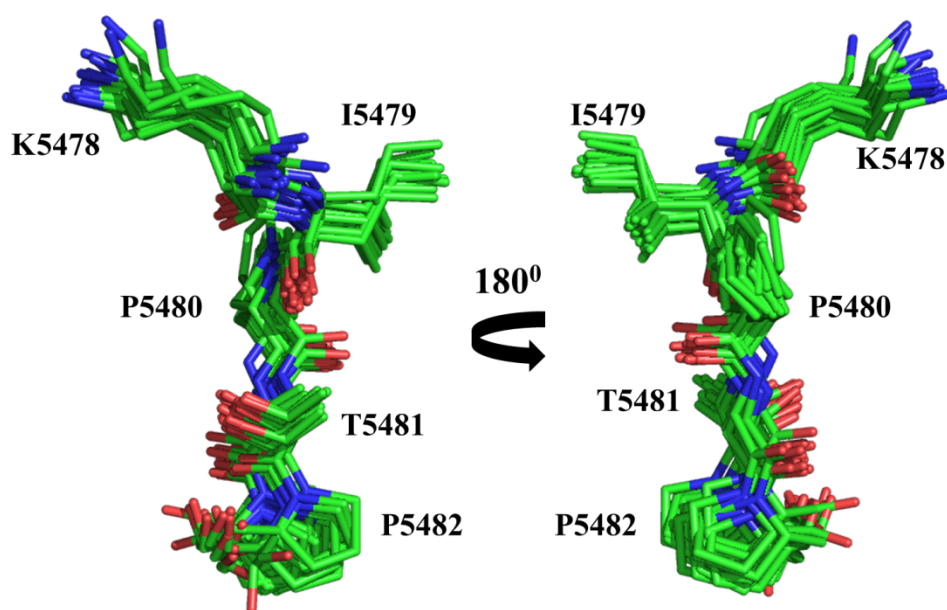


Figure 5.26 – Superimposition of the ensemble of structures obtained for the peptide in complex with EB1cΔ8

Regarding the contacts between the peptide and the helical region of EB1cΔ8, small number of restraints – eight in total, are observed for Lys5475 and Pro5476 and residues Asn223, Glu225 and Leu226. Ser5477 only shows one distance restraint to Phe218 and this is quite large distance  $\sim 6.5$  Å. The following residue, Lys5478 is solvent exposed making contacts only with the C-terminus of EB1cΔ8. Ile5479 shows the largest number of NOE distance restraints; these include the following residues from the coiled coil region of EB1cΔ8 – Tyr217, Phe218, Leu221, Arg222, Ile224, Glu225, Leu246 and Tyr247. Similarly to Ile5479, Pro5480 shows hydrophobic contacts to Tyr217, Phe218 and further down with Thr249. The following threonine (5481) is the only residue showing contacts to Phe216, making two additional contacts with other two aromatic side chains (Tyr217 and Phe218). Pro5482 is the last residue of the peptide making contacts with the coiled region of EB1cΔ8 – Thr249, already in the transition region to the C-terminal tail.

The above contacts from the NMR structures are generally consistent to the ones described by Honnappa and co-workers and the structure is very similar to the crystal structure 3GJO<sup>31</sup>. However, the NMR structure gives a better understanding of the solution structure of the complex, reassuring that the C-terminus folds upon binding and that this region seemed to be fixed in the crystal structure is not derived from



crystal packing. It is now very clear that the C-terminal plays a fundamental role in ligand-EB1c interactions. It was, therefore, decided to make an attempt in finding a higher affinity ligand by introducing mutations that can favour the interaction between the ligand and the C-terminus. This process will be subsequently described.

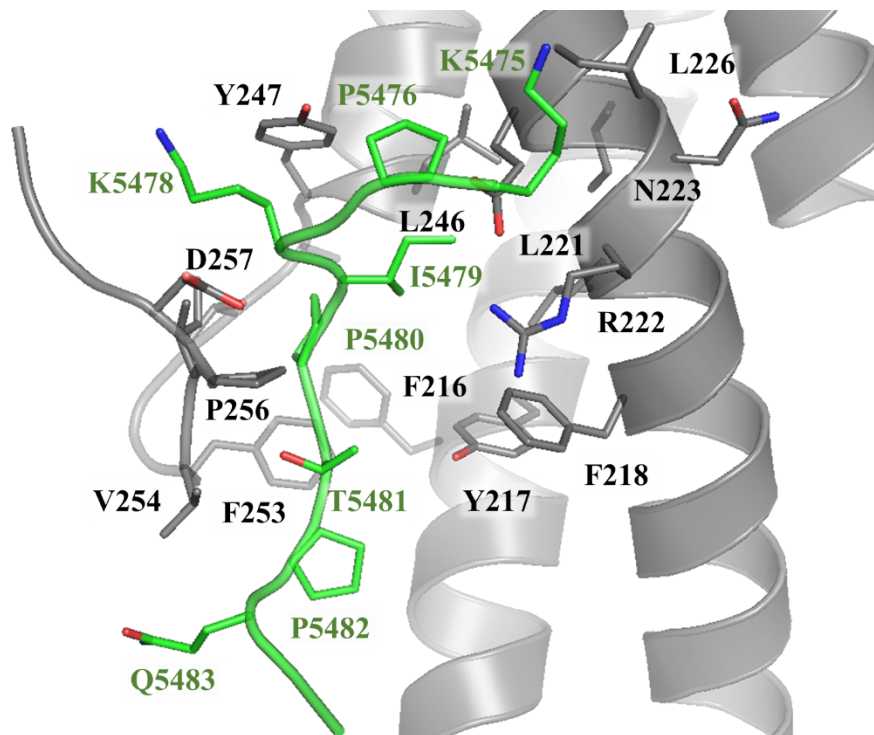


Figure 5.27 – representation of the residues forming the contact interface between EB1c $\Delta$ 8 and 11MACF, both shown in cartoon representation with side chain showed for all the residues involved in contacts between the two molecules. EB1c $\Delta$ 8 is shown in grey and 11MACF in green, with oxygen shown in red and nitrogen shown in blue for both.

## 5.5 Optimisation of the native interaction

### 5.5.1 Molecular modelling

At this stage it was known the importance of targeting the flexible C-terminus of EB1c, therefore it was decided to find ligands that bind tighter to this region through peptide mutagenesis. The first step consisted in a literature search to identify possible amino acid replacements that would help to achieve a stronger interaction. A paper published in 2012 by Buey et al., 2012<sup>49</sup> shows through a SPOT analysis that <sup>5481</sup>TPQ<sup>5483</sup> would tolerate mutations, especially if amino acid residues with hydrophobic side chains are introduced. Thr5481 can be replaced by histidine, isoleucine, leucine, methionine,

glutamine, arginine or valine. The following proline (Pro5482) tolerates phenylalanine and leucine, whereas the Gln5483 can be replaced by isoleucine, lysine, leucine or valine, being leucine the preferable mutation.

Three versions of mutated peptides were built and analysed through molecular docking – KPSKIPLLLRK (11MACF-LLL), KPSKIPVLLRK (11MACF-VLL) and KPSKIPILLRK (11MACF-ILL). Each pose was initially docked and evaluated using Goldscore scoring function (as described in section 8.1.2), and the results were later rescored using each scoring function present in GOLD – Chemscore, ChemPLP and ASP. Overall 11MACF-LLL obtained higher fitness scores for all scoring functions, followed by 11MACF-VLL. 11MACF (wild-type) and 11MACF-ILL seemed to compete for the third best overall score. These results are detailed in Table 5.10, Table 5.11, Table 5.12 and Table 5.13.

Table 5.10 – Fitness scores obtained with Goldscore for the wild-type and mutated versions.

	11MACF-wt	11MACF-LLL	11MACF-VLL	11MACF-ILL
1	116.6972	99.0181	109.2072	147.3196
2	95.596	98.691	103.7247	106.0801
3	90.5117	97.5353	88.4816	103.5024
4	86.0151	97.487	77.7022	102.2641
5	75.067	95.2011	76.4867	95.2081
6	68.4973	89.3229	44.6624	78.6898
7	67.8119	84.465	42.4577	34.2308
8	50.9331	57.7077	41.1628	22.3123
9	-20.8909	52.4758	14.3184	-70.6193
10	-34.9973	2.4104	8.3302	-241.4937
average	59.5241	77.4314	60.6534	37.7494

Table 5.11 – Fitness scores obtained with Chemscore for the wild-type and mutated versions.

	11MACF-wt	11MACF-LLL	11MACF-VLL	11MACF-ILL
1	-22.7581	-9.5765	-17.2281	-5.7635
2	-46.008	-19.7418	-26.6175	-17.1386
3	-22.0417	-19.0367	-7.6088	-26.7014
4	-45.7151	-11.024	-21.107	-11.639
5	-41.8411	-13.8251	-44.2843	-12.9866
6	-41.8287	-17.7299	-29.9582	-29.5592
7	-43.758	-15.5714	-26.3761	-50.1637
8	-38.263	-23.1886	-41.2771	-51.7802
9	-44.3287	-29.5566	-47.0179	-55.7457
10	-59.8753	-29.5404	-29.8831	-93.4951
average	-40.642	-18.879	-29.136	-35.497

Table 5.12 - Fitness scores obtained with ChemPLP for the wild-type and mutated versions.

	11MACF-wt	11MACF-LLL	11MACF-VLL	11MACF-ILL
1	103.8213	-405.4915	94.0879	89.2584
2	43.1355	90.8319	-1055.4376	95.9891
3	88.657	91.1274	82.2654	-8942.0868
4	55.8481	97.2031	75.0087	109.75
5	-2938.9568	-10005.363	-465.8373	-979.9262
6	32.0876	15.2989	35.0926	79.6759
7	73.1813	88.985	65.077	29.6528
8	-2977.9218	-539.0405	41.4024	26.4023
9	72.9775	73.5982	-8.4312	21.6552
10	17.0947	-8454.1646	65.6476	-63.757
average	-543.0076	-1894.702	-107.1125	-953.3386

Table 5.13 - Fitness scores obtained with ASP for the wild-type and mutated versions.

	11MACF-wt	11MACF-LLL	11MACF-VLL	11MACF-ILL
1	37.9338	19.4951	29.0991	30.3398
2	6.7662	31.8461	28.7736	35.4951
3	31.473	31.8952	24.2367	26.3963
4	-15.557	34.8758	27.5771	23.7509
5	28.6352	26.0728	18.879	18.5394
6	17.517	33.6309	2.3211	22.9137
7	15.0886	11.4932	22.6155	-23.579
8	-0.3545	11.6971	-3.3258	-9.5648
9	-12.055	18.9143	11.5336	-30.678
10	3.7372	17.2829	10.1022	-49.704
average	11.319	23.72	17.181	4.391

Regarding the best scored pose for each peptide the most relevant differences that can be observed for the mutated region are graphically shown in Figure 5.28. The wild-type peptide shows two hydrogen bond contacts between Thr5481 (T7) backbone and Val254 backbone. Pro5480 (P6) carbonyl makes an extra hydrogen bond with the OH group of Tyr217 and Pro5482 (P8) is involved in hydrophobic interactions with Gly252 and Phe253.

11MACF-LLL does not form the hydrogen bond previously observed between the threonine backbone and Val254. Instead Lys5478 (K4) establishes a hydrogen bond with Pro256 and the terminal Lys (K11) establishes a similar contact with Ile255. Despite loss of two hydrogen bonds with the valine two other contacts were formed with the two subsequent residues, Ile255 and Pro256. In terms of hydrophobic interactions Val254 is now accessible to make contacts with both side chains of Leu5482 (L8) and Lys5485 (K11). Leu 5482 (L8) can also make hydrophobic contacts with the hydrophobic Phe253 ring and Leu5481 (L7) is accessible to make hydrophobic contact with Phe218. The docking prediction reveals that this mutant is likely to give extra hydrophobic contributions on EB1 binding, as it could be expected due to its hydrophobic nature.

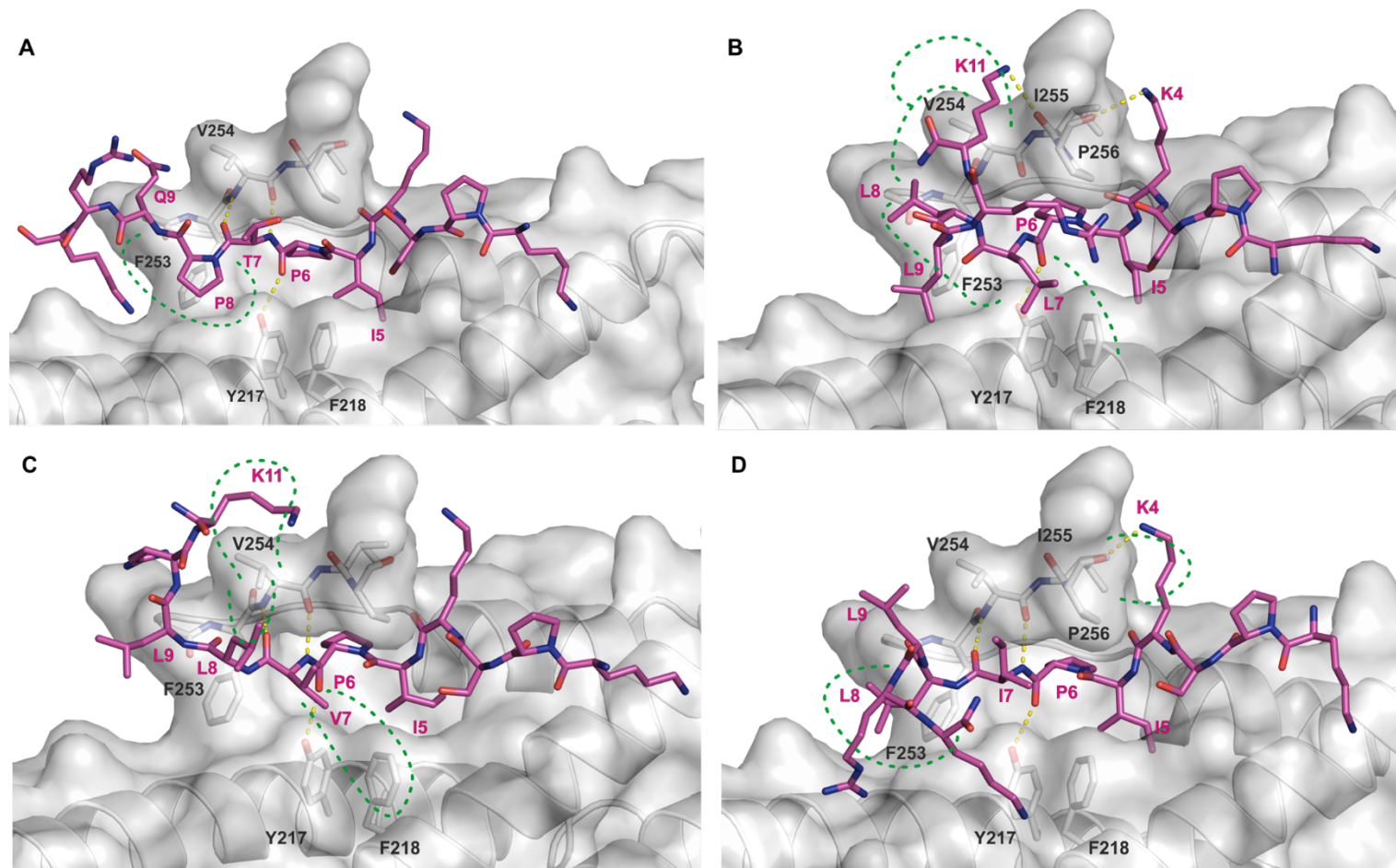


Figure 5.28 – Best scored docking poses obtained for A – 11MACF-wt, B – 11MACF-LLL, C - 11MACF-VLL and D – 11MACF-ILL. Peptides carbon atoms are coloured in magenta, and EB1's in light grey. Oxygen is shown in red and nitrogen in blue. Hydrogen bonds are shown as yellow dashed lines and hydrophobic interactions regions are identified by a green dashed line.

The replacement of Thr5481 by a valine instead of a leucine – 11MACF-VLL, does not affect the accessibility of the backbone to establish hydrogen bond contacts with Val254. The reason seems to be the size of the side chain. Whereas leucine is a  $\gamma$  branched hydrophobic side chain, valine is a  $\beta$  branched hydrophobic side chain similarly to threonine. This difference does not affect the spatial rearrangement of the peptide side chains in such a way that the backbone of position 7 is not accessible to establish hydrogen bond contacts with the backbone of Val254. In the same manner the hydrogen bond with the OH of Tyr217 is retained. However, and as expected, extra hydrophobic contributions are achieved in comparison with the wild type peptide. The side chain of V7 seems to be long enough to establish hydrophobic contacts with the side chain of Phe218, as previously observed for the leucine. Similarly to 11MACF-LLL the side chain of Val254 seems to make hydrophobic contacts with both side chains of L8 and K11.

Finally, 11MACF-ILL shows the same pattern of hydrogen bonds observed for both wild type and 11MACF-VLL (Val254 and Tyr217). The most prominent hydrophobic contacts between this peptide and the C-terminus of EB1 are between L8 and Phe253 and K4 and Ile255.

Overall and based on both fitness scores and predicted binding poses it was decided to test two peptides – 11MACF-LLL and 11MACF-VLL. The biophysical assays results obtained will be subsequently described in the next section.

### **5.5.2 NMR screening of MACF mutated peptides against EB1 $\Delta$ 8**

The NMR titrations of 11MACF-LLL and 11MACF-VLL show the chemical shifts changes of the NH resonances are in slow exchange regime and therefore the complexes formed are tightly bound - Figure 5.29 and Figure 5.30.

When 12.5  $\mu$ M of ligand is added to EB1 $\Delta$ 8, the intensity of the peaks of the protein in the free form start to decrease in intensity. Between 50-75  $\mu$ M of ligand a peak correspondent to the bound form appears in a distinct position, and the intensity increases with ligand addition reaching saturation between around 1:1 ligand:protein ratio (100  $\mu$ M) - Figure 5.31.

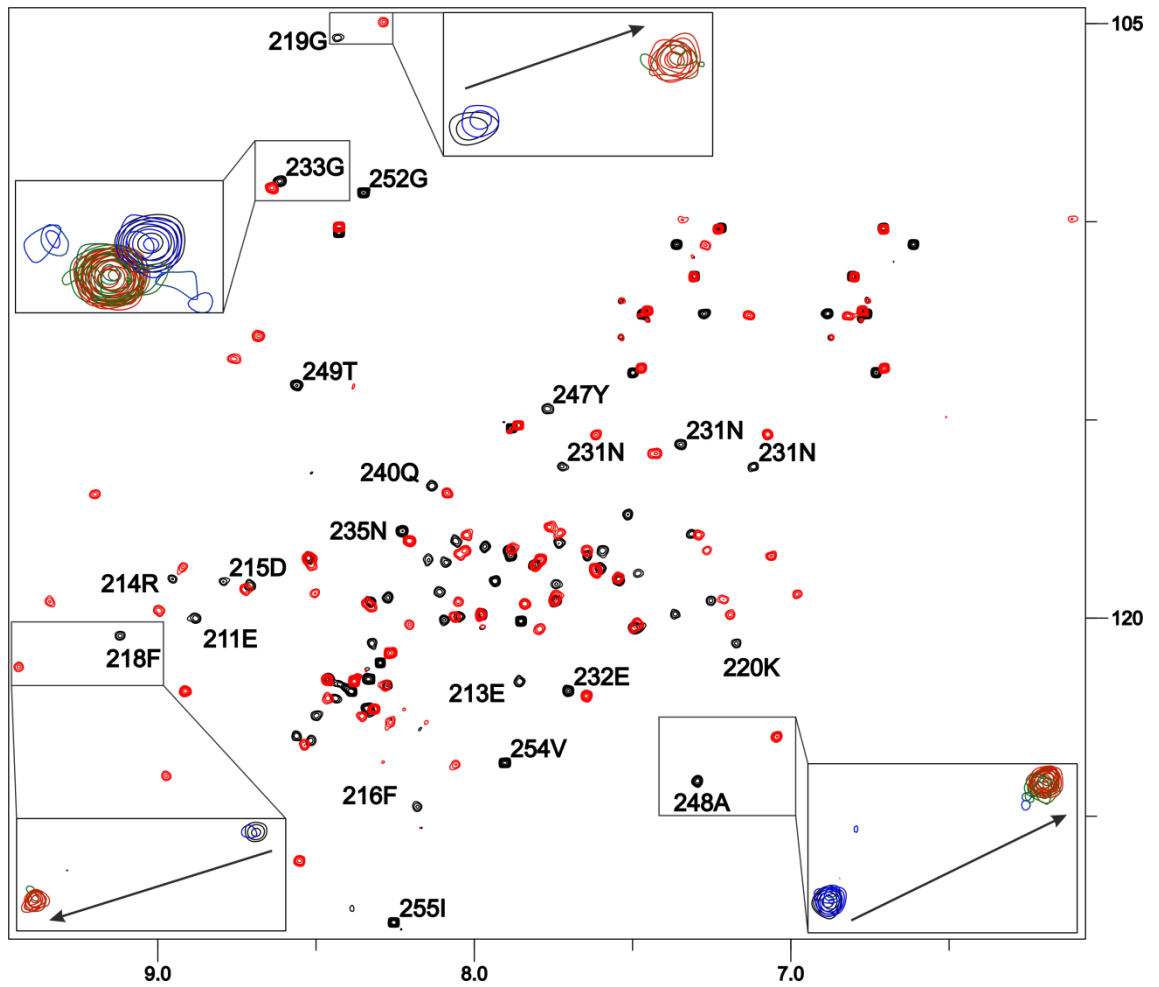


Figure 5.29 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 8 (100  $\mu\text{M}$ ) recorded at 600 MHz - in the free form (black) and in the presence recorded of 11MACF-LLL (400  $\mu\text{M}$ , red). The insets show regions of interest and the spectra corresponding to the following 11MACF-LLL concentrations, 12.5, 25, 75, 82.5, 100, 200 and 400  $\mu\text{M}$ .

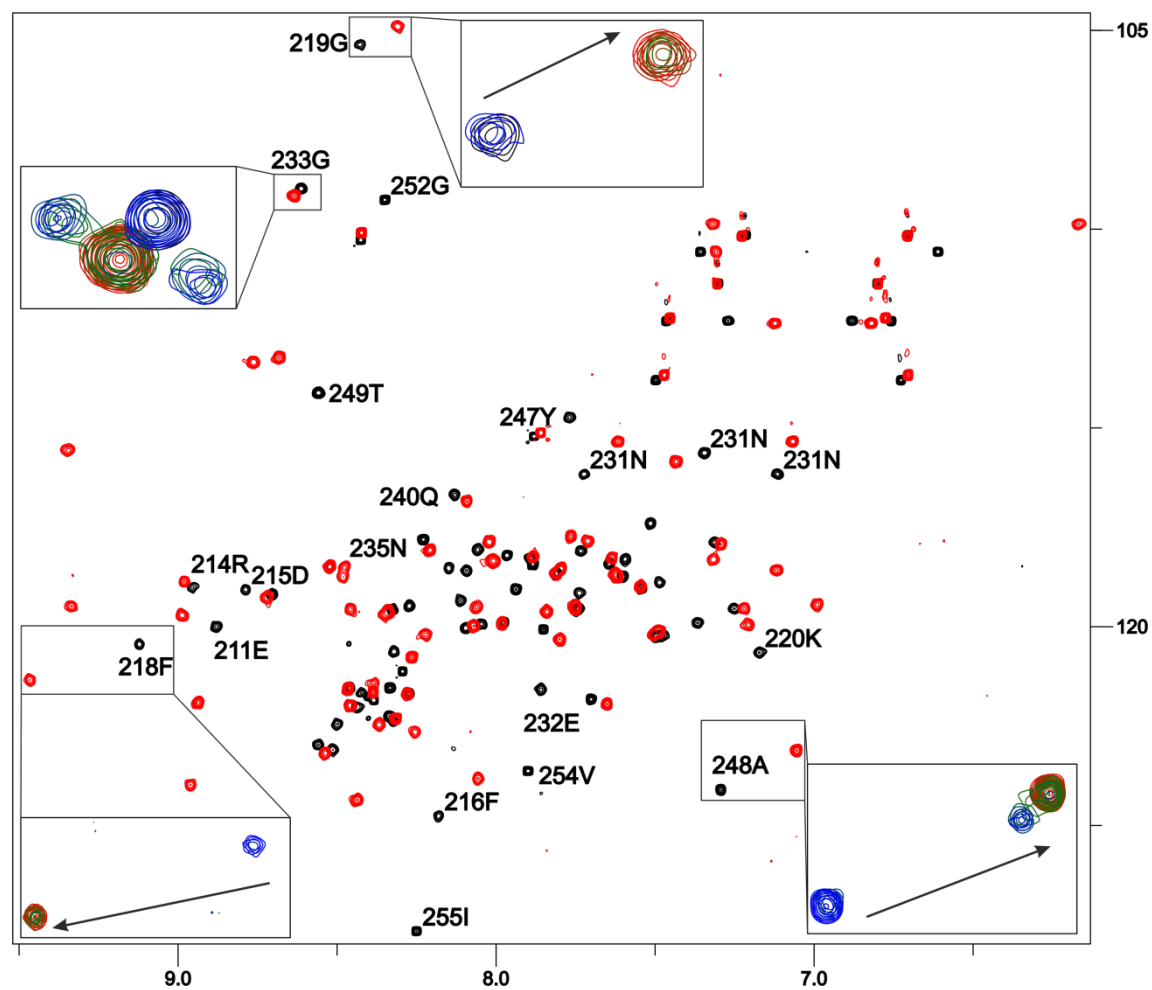


Figure 5.30 - Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labelled EB1c $\Delta$ 8 (100  $\mu\text{M}$ ) recorded at 800 MHz - in the free form (black) and in the presence recorded of 11MACF-VLL (400  $\mu\text{M}$ , red). The insets show regions of interest and the spectra corresponding to the following 11MACF-VLL concentrations, 12.5, 25, 75, 82.5, 100 and 200  $\mu\text{M}$ .



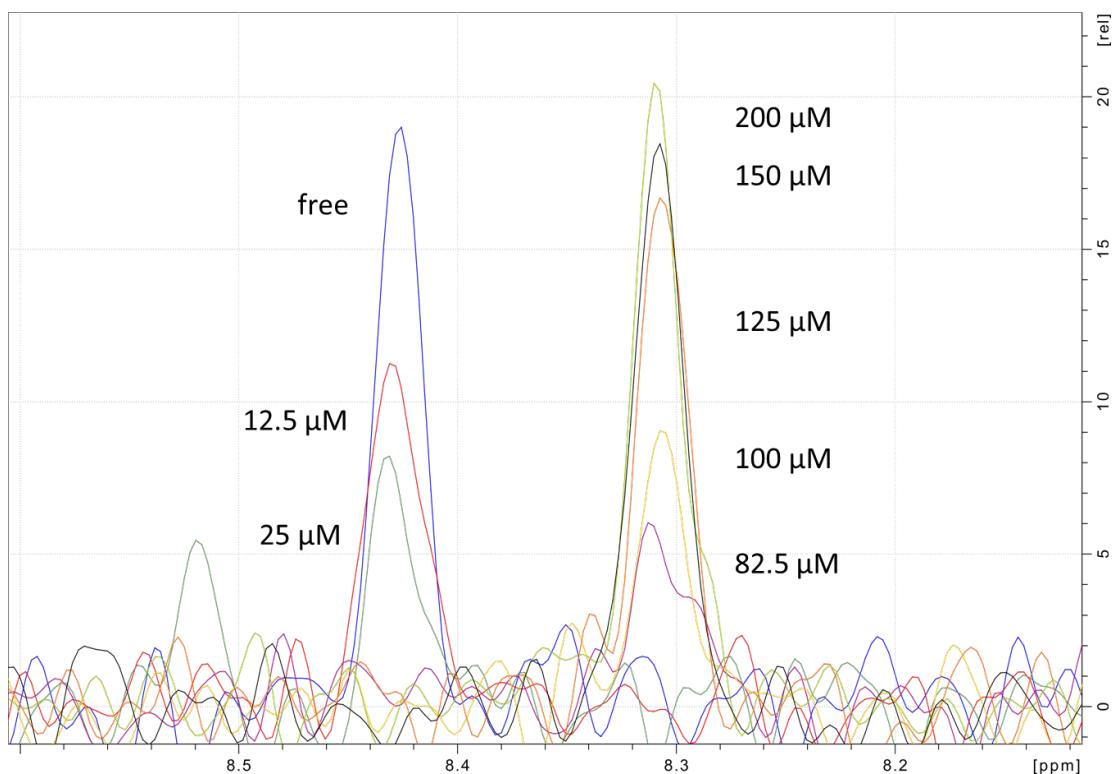


Figure 5.31 – superimposition of the proton projection of Gly219 from  $^1\text{H}, ^{15}\text{N}$ -HSQC titration experiments performed for EB1cΔ8-11MACF-VLL. On the left hand side one can observe the decrease in intensity of the peak correspondent to the free form and sequential increase of intensity of the peak for the resonance of the complex (right hand side).

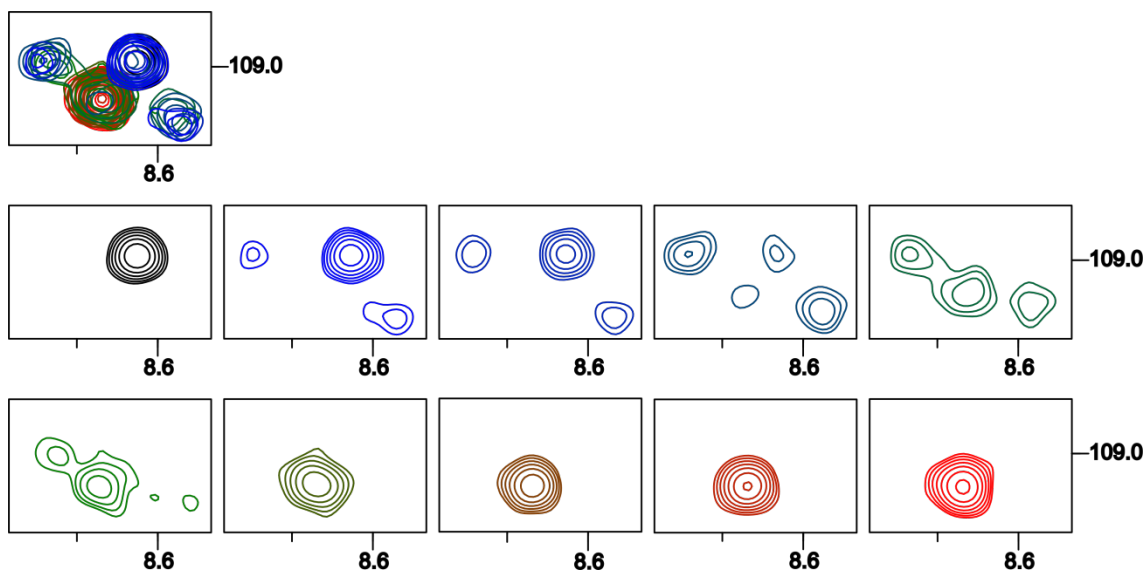


Figure 5.32 – Evolution of the peak changes for Gly233 of EB1cΔ8 upon titration with 11MACF-VLL. The top square corresponds to the superimposition of all titration points, and below each titration point represented individually. The peaks are coloured in the following manner: free protein – black, followed by 12.5 (navy), 25 (blue), 50 (teal), 75 (olive), 82.5 (green), 100 (khaki), 125 (brown), 150 (orange) and finally 200  $\mu\text{M}$  of ligand – red.

Interestingly, for some residues in the four bundle helix region such as Asn231 and Gly233, four resonances are observed in the intermediate titration points (ligand concentration 12.5 – 82.5  $\mu\text{M}$ ) - Figure 5.32. Because EB1c is a homodimer and has two symmetric binding sites, it is possible that the extra two peaks observed before the complex reaches saturation belong to the conformations of EB1c $\Delta$ 8 with only one of the binding sites occupied by the ligand.

Similarly to what was observed for 11MACF-wild type, the Val254 NH cross peak for both mutants, 11MACF-LLL and VLL, is the most affected upon ligand binding –  $\Delta\delta = 2.9$  ppm and  $\Delta\delta = 3.4$  ppm, respectively. Overall the chemical shift changes follow a similar pattern to the one obtained for the wild-type, with significant chemical shift changes in the C-terminus for Asp250 ( $\Delta\delta = 1.8$  and 1.7 ppm), Gly252 ( $\Delta\delta = 1.7$  and 1.6 ppm) and Ile255 ( $\Delta\delta = 0.7$  and 1.3 ppm), around the binding site for the SxIP – Phe218 ( $\Delta\delta = 0.4$  and 0.5 ppm), Lys220 ( $\Delta\delta = 0.5$  ppm), Leu221 ( $\Delta\delta = 0.3$  ppm), Glu225 ( $\Delta\delta = 0.6$  ppm), Leu246 ( $\Delta\delta = 0.3$  ppm), Tyr247 ( $\Delta\delta = 1.7$  ppm), Ala248 ( $\Delta\delta = 0.5$  ppm), Thr249 ( $\Delta\delta = 0.5$  and 0.4 ppm), immediately below – Glu213 ( $\Delta\delta = 0.5$  ppm) or immediately above – Cys228 ( $\Delta\delta = 0.4$  ppm).

In terms of total distance observed for the chemical shift changes the general trend is 11MACF-VLL promotes larger chemical shift changes. Good examples of this trend are Glu213, Arg214, Asp215, Phe216, Phe218, Arg222, Tyr247 and even more noticeable Val254 and other residues from the C-terminus such as Phe253 and Ile255. However, this is not true for all resonances, with Tyr217, Lys220 and Thr249 being examples where the chemical shift changes are larger for the wild-type peptide.

The shape of the chemical shift changes also change from a complete linear evolution – Phe218, Ala248, to a curved shape – Phe216, Gly219 and Lys220 (Figure 5.33). Whereas a linear trend usually means that the binding mode between ligands is the same just different strength, the curvature indicates there are differences in binding mode between all three ligands. Val254 is an interesting case with the chemical shift change being very large for the wild-type and progressing in a curved manner for the mutated peptides. This may mean this residue is affected in a different manner depending on the ligand, which is something to expect since the designed peptides are expected to interact with residue in different ways – 11MACF-LLL maximises the hydrophobic contacts and 11MACF-VLL should have the hydrophobic contacts enhanced similarly to the first mutant, but keep the hydrogen bonds present for the

wild-type.

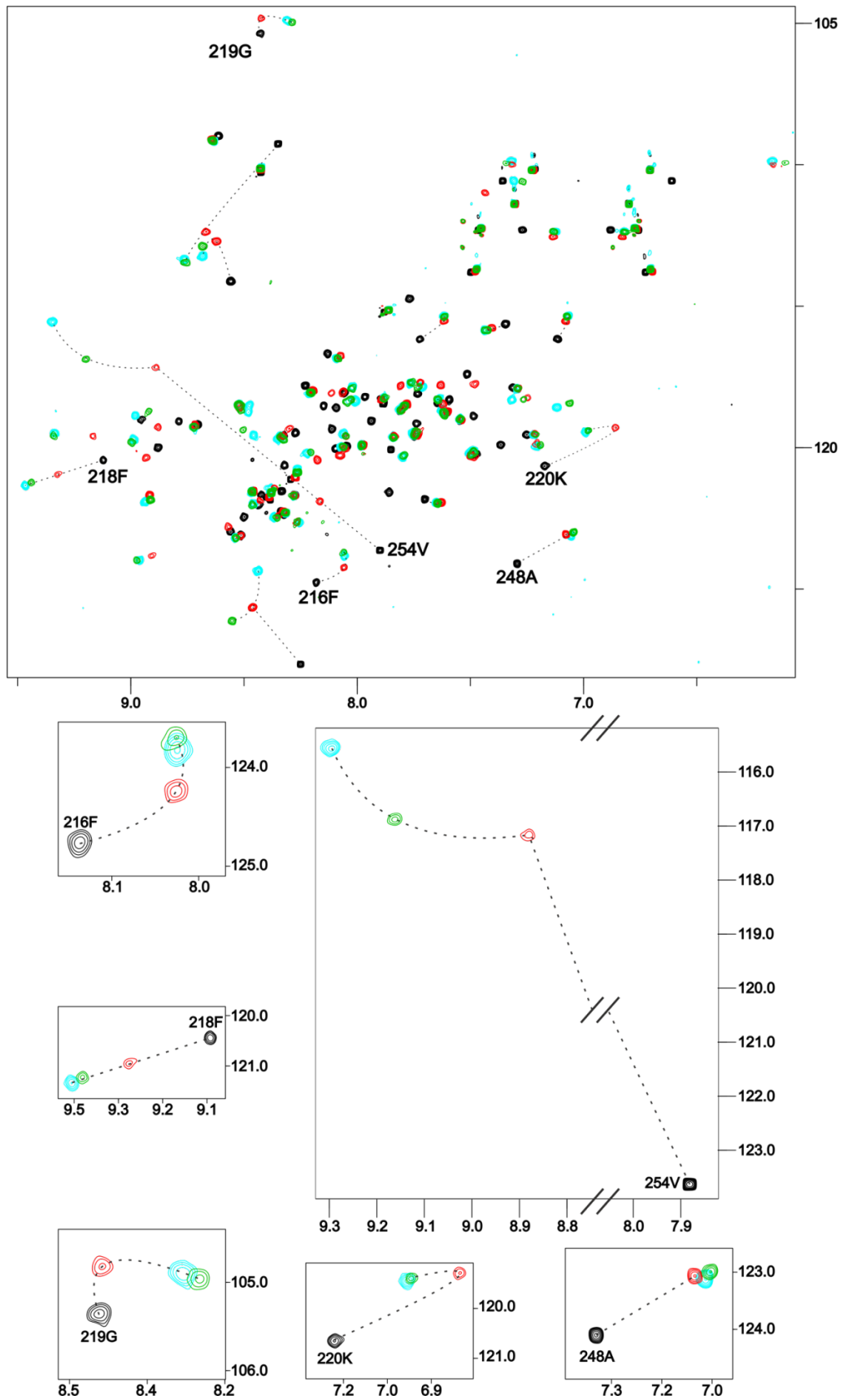


Figure 5.33 – Superimposition of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC for free EB1c $\Delta$ 8 (black) and final titration points for

11MACF-wt (red), 11MACF-LLL (green) and 11MACF-VLL (cyan). The dotted lines represent the trajectory of the chemical shift.

### 5.5.3 Isothermal titration calorimetry binding assays of MACF mutated peptides

The NMR titrations indicate that both 11MACF-LLL and 11MACF-VLL bind tighter to EB1c $\Delta$ 8 than the wild type 11MACF, indicating that the introduction of amino acid residues with hydrophobic side chains was beneficial to the binding. ITC can provide important information that can help to further understand the binding mechanism through determination of entropic and enthalpic contributions for the binding.

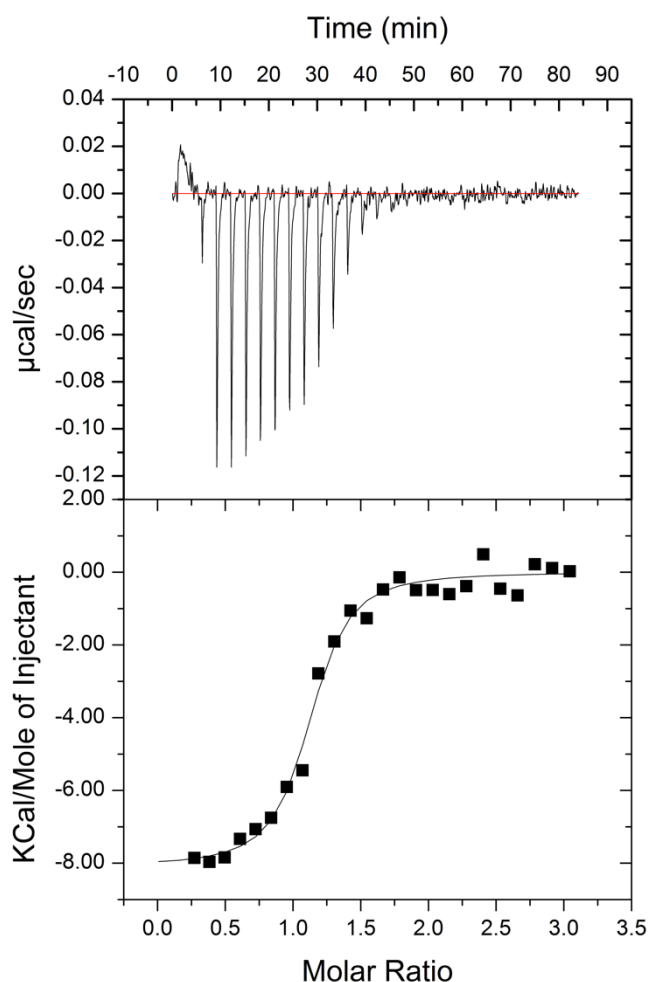


Figure 5.34 - The ITC isotherm (top) and resultant curve (bottom) produced by the ITC binding experiment between 15  $\mu\text{M}$  EB1c $\Delta$ 8 (cell) and 225  $\mu\text{M}$  11MACF-LLL (syringe) in 20 mM phosphate, pH 6.5, 50 mM NaCl, 0.5  $\mu\text{M}$  TCEP, 0.02% NaN<sub>3</sub> on an iTC<sub>200</sub> Microcalorimeter (MicroCal) at 25°C. Fitting of the curve produced by the ITC experiment to a single set of sites curve-fitting model, using Origin7, resulted in a binding affinity ( $K_d$ ) of 340 nM.

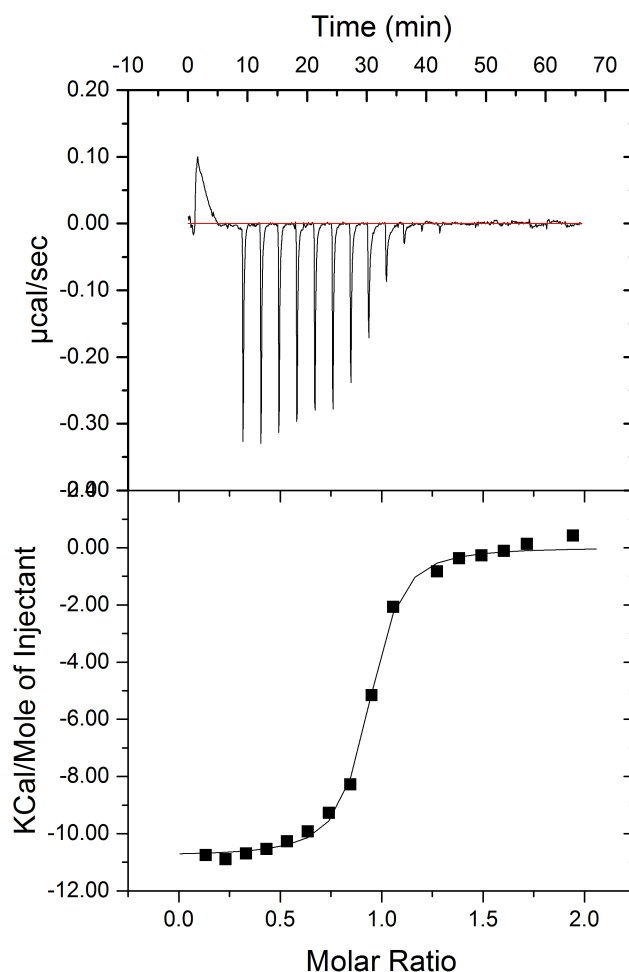


Figure 5.35 - The ITC isotherm (top) and resultant curve (bottom) produced by the ITC binding experiment between 25  $\mu\text{M}$  EB1c $\Delta$ 8 (cell) and 250  $\mu\text{M}$  11MACF-VLL (syringe) in 20 mM phosphate, pH 6.5, 50 mM NaCl, 0.5  $\mu\text{M}$  TCEP, 0.02%  $\text{NaN}_3$  on an iTC<sub>200</sub> Microcalorimeter (MicroCal) at 25°C. Fitting of the curve produced by the ITC experiment to a single set of sites curve-fitting model, using Origin7, resulted in a binding affinity ( $K_d$ ) of 163 nM.

Table 5.14 - Thermodynamic parameters obtained for three ITC binding experiments to detect the interactions of 11-residue peptides with EB1c $\Delta$ 8. Values shown in this table correspond to the average and associated error obtained for the three experiments.

Interaction	$K_d$ ( $\mu\text{M}$ )	$\Delta H$ ( $\text{kcal mol}^{-1}$ )	$-T\Delta S$ ( $\text{kcal mol}^{-1}$ )	$\Delta G$ ( $\text{kcal mol}^{-1}$ )
EB1c $\Delta$ 8-11MACF-wt	$3.80 \pm 0.82$	$-10.57 \pm 0.63$	$3.15 \pm 0.64$	$-7.42 \pm 0.12$
EB1c $\Delta$ 8-11MACF-LLL	$0.030 \pm 0.004$	$-8.26 \pm 0.13$	$-0.65 \pm 0.16$	$-8.91 \pm 0.13$
EB1c $\Delta$ 8-11MACF-VLL	$0.015 \pm 0.001$	$-10.50 \pm 0.21$	$1.19 \pm 0.21$	$-9.30 \pm 0.03$

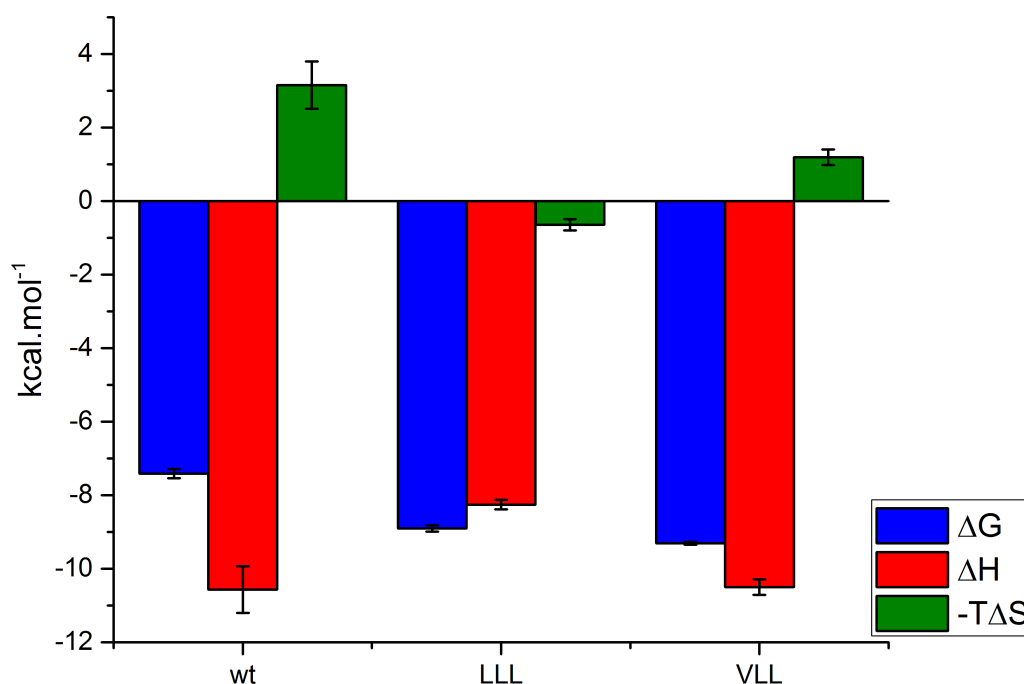


Figure 5.36 - The thermodynamic parameters ( $\Delta G$  (blue),  $\Delta H$  (red) and  $-T\Delta S$  (green)) obtained from the binding interaction between the 11MACF-wt (left hand side), 11MACF-LLL (middle), 11MACF-VLL (right hand side) and EB1c $\Delta 8$ .

The replacement of the TPQ sequence in the wild type peptide by a leucine patch (LLL) yielded a binding affinity increase of 10 fold ( $K_d$  changed from 3 to 0.3  $\mu\text{M}$ ) – Figure 5.34, Table 5.14 with  $\Delta G$  value changing from -7.42 to -8.91 kcal mol<sup>-1</sup>. Due to the very significant improvement in the  $K_d$  value, one perhaps could expect a greater affinity in the overall energy of the reaction ( $\Delta G$ ). However, the enthalpic contributions for the binding are now smaller than for the wild type, penalising  $\Delta G$ . A possible explanation can be the unfavourable position of the backbone of position 7 of the peptide to establish hydrogen bonds with the backbone of Val254 of the C-terminus – Figure 5.28. Entropic contributions underwent a notable improvement and these can be associated with the greater affinity of this interaction when compared with the wild type ligand – Figure 5.36. Again, molecular docking can justify partially these results. Figure 5.28, shows that this the residue in position 7, leucine, can now establish hydrophobic interactions with Phe218. In addition, hydrophobic contributions from the two subsequent leucine residues with Val254 and Phe253 may justify the higher entropic contribution for the formation of this complex (section 5.5.1).

For the second peptide, 11MACF-VLL, a similar trend is observed, the introduction of hydrophobic side chains have a beneficial contribution for the binding. The  $K_d$  is

now about half that the one observed for 11MACF-LLL, meaning it binds about two times tighter – Table 5.14.  $\Delta G$  is now  $-9.30 \text{ kcal mol}^{-1}$  and the lowest for the three tested peptides. The enthalpic contribution improved regarding the complex with 11MACF-LLL and it is now very close to the wild type. In terms of entropic contributions, this complex shows an improvement compared with the wild type ( $1.19 \text{ kcal mol}^{-1}$  compared with  $3.15 \text{ kcal mol}^{-1}$ ) but not as good as 11MACF-LLL ( $-0.65 \text{ kcal mol}^{-1}$ ). The size of the side chains of threonine and valine is similar, with the advantage of the later can provide extra hydrophobic interactions. It is still possible the existence of hydrophobic interaction between the valine in position 7 and Phe218. However due to the shorter length of valine side chain, this may not be as strong as the one provided by the leucine side chain. Hydrophobic interactions with Val254 are still possible, although the interaction with Phe253 seems to have been lost.

The enthalpy/entropy compensation is a common issue and invariably accompanies protein-ligand associations, improving binding enthalpy or entropy changes does not guarantee enhanced binding affinity because gains in one thermodynamic parameter may be offset by losses in the other.<sup>16></sup> With 11MACF-VLL a good balance between both entropic and enthalpic terms was achieved.

Usually establishing strong interactions between polar groups on the ligand and the protein is difficult in practice because the energetics of these interactions (hydrogen bonds) are strongly distance and angle dependent. Non-directional van der Waals contacts are less directional, but dispersion forces are less specific and weaker than polar interactions.<sup>16></sup>

The contribution of the solvent to the enthalpic contributions brings extra complication to the analysis of thermodynamic contributions in a protein-ligand reaction. Another difficulty in optimising bimolecular associations of a protein and ligands using thermodynamic information analysis is the loss in rotational and translational entropy upon binding may have effects on the entropic term of the binding event. Therefore analysis of the influence of thermodynamic parameters and their effects on the binding is not a straightforward relationship.<sup>16></sup>

## **5.6 Elucidation of the binding contributions to EB1c and the role of the C-terminus**

The results described in this chapter demonstrate clearly the importance of the C-terminus tail in EB1c binding. In this section, the contributions from different regions of the protein and ligands for the interaction will be analysed in order to understand fully how EB1 recruits SxIP proteins and how that knowledge can be used to design higher affinity ligands.

The short consensus sequence, SKIP, is sufficient and the key consensus sequence to specifically interact with EB1c's hydrophobic cavity. Unfortunately, the binding affinity is very weak (10 mM) and consequently not enough to achieve a biological response – binding affinities reported for MACF and APC derived peptides (~30 residues) of 3.5 and 5.8  $\mu\text{M}$  respectively.<sup>31,49</sup> Increasing the ligand length by two residues, SKIPTP, improves the binding constant to about six fold without losing the specific binding. The two additional residues, proline and threonine, seem to engage with the lower region of the four helix bundle, just below the SxIP binding site, with NOEs to the aromatic patch <sup>216</sup>FYF<sup>218</sup>, and contacts with the C-terminus – <sup>253</sup>FVIP<sup>256</sup>. This was expected as the crystal structure reported by Honnappa and co-workers in 2009<sup>31</sup> described hydrogen bond formation between Val254's backbone and Thr5481, indicating this residue could have an important role in the interaction. Interaction between this peptide (SKIPTP) and EB1c $\Delta$ 8 seem to have a fair contribution from the C-terminus as the  $\Delta\text{G}$  increases from -3.8 kcal.mol<sup>-1</sup> to -2.9 kcal.mol<sup>-1</sup> (0.9 kcal.mol<sup>-1</sup>) upon removal of the C-terminal tail of EB1c. Regarding the 11 residue peptide, with addition of two amino acids in the N-terminus (KP) and three in the C-terminus (QRK), the  $\Delta\text{G}$  contribution almost doubles the one observed for the six residue peptide. Removal of the binding contributions of the C-terminus have now a larger effect in the  $\Delta\text{G}$ , accounted as 1.42 kcal.mol<sup>-1</sup>, instead of 0.9 kcal.mol<sup>-1</sup>. In summary, the contributions of the peptide residues for the binding to EB1c $\Delta$ 8 can be observed in Figure 5.37, where the SKIP sequence has a contribution of -2.73 kcal.mol<sup>-1</sup>, the TP give an extra contribution of -1.05 kcal.mol<sup>-1</sup> (where -0.2 kcal.mol<sup>-1</sup> come from interactions with the C-terminus of EB1c $\Delta$ 8). Finally, the full 11 residue native sequence adds an extra -3.64 kcal.mol<sup>-1</sup> where more than half, 2.22 kcal.mol<sup>-1</sup> comes from interactions with the C-terminus.



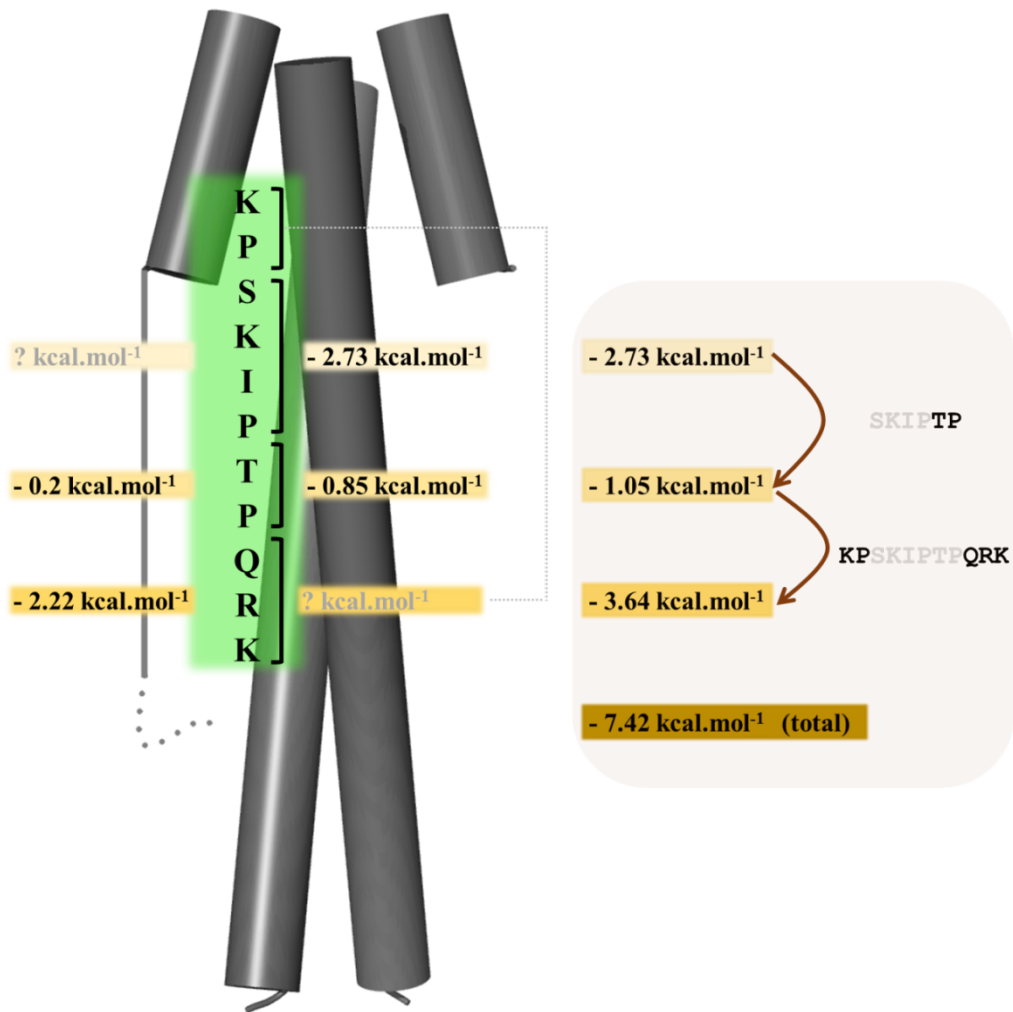


Figure 5.37 – schematic representation of the contributions expressed in  $\Delta G$  for the binding of a native sequence peptide derived from MACF2 to EB1c $\Delta$ 8. Peptide corresponds to the green box and the protein is represented in grey, the C-terminal tail contributions are shown on top of the schematic representation of the C-terminus, and contributions independent of this region of EB1c are shown immediately on the right hand side of the peptide. Finally, in the box on the right hand side there is a representation of the total contributions *per* peptide length.

Concerning the peptides designed to improve the contributions with the C-terminus of EB1c and therefore have higher affinities to EB1c, only one of the mutants, 11MACF-LLL, was tested in the absence of the C-terminus (EB1c $\Delta$ 16). The difference in  $\Delta G$  is even larger than the one observed for the native sequence, 2.31 kcal.mol<sup>-1</sup> compared with the previous 1.42 kcal.mol<sup>-1</sup>. This was expected as this peptide was designed to improve the hydrophobic interactions with the C-terminus and removing this region of the protein has a more pronounced effect in the binding when compared with other ligands. The last mutant to be tested, 11MACF-VLL, is the best ligand tested in the course of this project with a  $K_d$  of 150 nM and a  $\Delta G$  of -9.3 kcal.mol<sup>-1</sup>, indicating that

hydrophobic side chains immediately after the SxIP motif seem to be beneficial to the binding to EB1c.

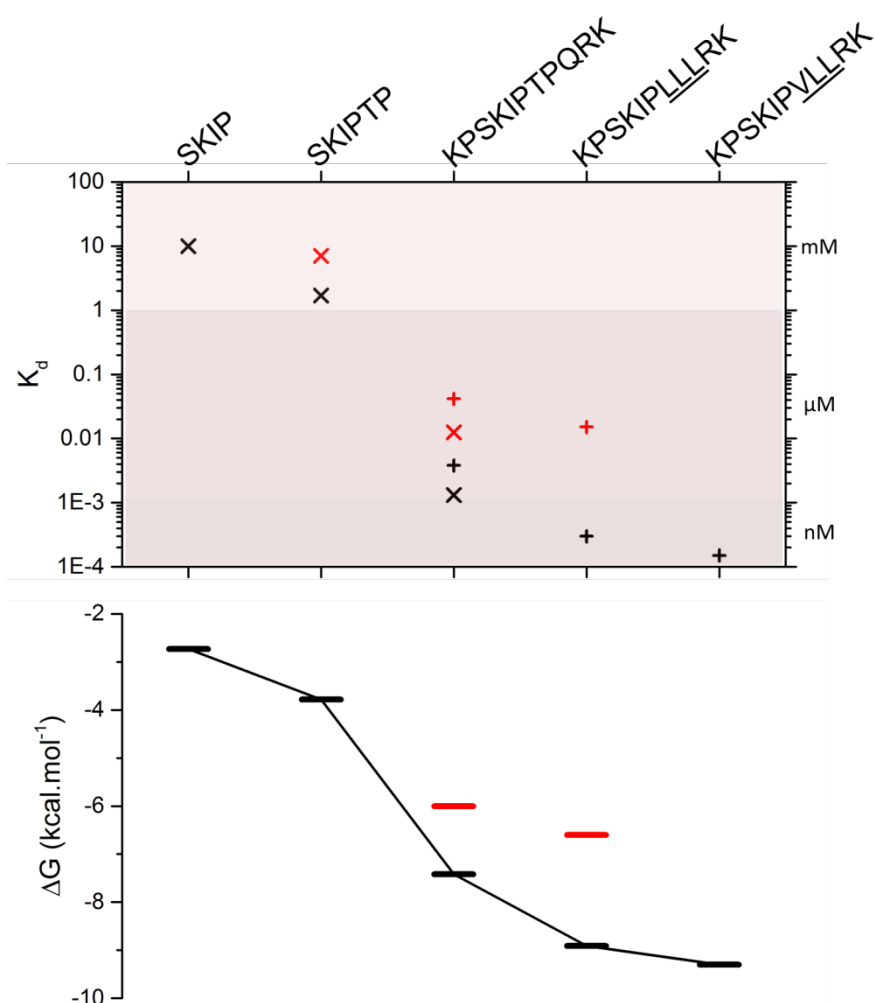


Figure 5.38 – summary of  $K_d$  and  $\Delta G$  values obtained using ITC and NMR methods for the tested peptides (wild-type and mutated) against EB1c with and without the C-terminus tail, EB1c $\Delta$ 8, (black) and EB1c $\Delta$ 16 (red), respectively. Top panel shows  $K_d$  values *per* peptide. NMR data is presented as a diagonal cross and ITC as a vertical cross. Bottom panel shows  $\Delta G$  values for each peptide.

The experiments conducted using different length peptides containing the SxIP motif and described in this chapter show that while the SxIP motif is sufficient to target EB1c, extra interactions, specifically with the flexible C-terminus are necessary to achieve higher affinity. This was further demonstrated by the higher affinity observed in peptides designed for enhanced interaction with the C-terminus – 11MACF-LLL and 11MACF-VLL. With such optimised ligands it is possible to build another pharmacophore model and use it to find better small molecule ligands to interact with EB1c $\Delta$ 8. For that, a three-dimensional structure model (X-ray or NMR) would ideally provide the information necessary for another iteration in small molecule identification.

Unfortunately, structural biology approaches to elucidate the three-dimensional structure of proteins are complicated processes and take long time. Therefore, this will be left outside the scope of this project and be considered for future work.

Next Chapter will cover a final discussion/conclusion on what was learned during the development of this project and what can be done in the future not only to target the EB1-SxIP interaction but also for protein-protein interactions in general.



## Chapter 6. CONCLUSIONS AND FUTURE WORK

For many years now the importance of the interactome—the complex network of protein-protein interactions—has been recognised as of great value for both biological systems and the development of disease states. Despite this, small-molecule drugs that act by directly disrupting or promoting the interaction between two proteins are relatively rare in comparison to other drug classes, and protein–protein interactions (PPIs) are viewed as challenging—in some cases essentially ‘undruggable’—targets.<sup>17</sup> Nevertheless, the last decade has seen amazing progress in tackling PPI targets with synthetic molecules. More than 40 PPIs have now been targeted and several inhibitors have reached clinical trials.<sup>87</sup>

The main aim of this project was to identify a chemical scaffold based on the SxIP motif that can target the EBH domain of EB1. For this purpose an initial virtual screening methodology was delineated based on the SxIP scaffold. The method developed used a multidisciplinary approach, with an initial pharmacophore model used for screening of a large virtual database, followed by molecular docking of the selected compounds and a final balanced ranking where docking scores and desirable molecular properties were taken into account. The approach proved to be successful as two molecules were found to bind specifically to the SxIP binding site. An advantage of this method is the fact that is highly targeted when compared with other methodologies such as high-throughput screening (HTS), meaning it is possible to obtain hit molecules in an inexpensive way, without needing large physical libraries of compounds, protein and developing a HTS assay.

Experimental data on the binding mode of the compounds to the protein by X-ray or NMR methods is often important and of great use to understand the mode of binding and is crucial for ligand optimisation.<sup>17</sup> Having two active hits with the same active scaffold prompted the structural elucidation of EB1c $\Delta$ 8 in the free form and in complex with one of these molecules. The data shows that contrary to what was initially assumed based on the X-ray structure of EB1c $\Delta$ 8 bound to 11MACFp1<sup>31</sup>, the binding pocket for the IP motif is dynamic in solution and may be completely absent. However, the IP mimetic scaffold identified in a previous chapter (Chapter 2) seems

to stabilise the two dynamic side chains, Arg222 and Tyr247 that affect the binding pocket size and conformation, open vs closed.

Despite the stabilisation of the binding pocket the compounds identified in Chapter 2 do not have high affinity for EB1cΔ8 with values of 6-10 mM, having slightly higher affinity than the *microtubule tip localization signal*<sup>31</sup> (SxIP), 14 mM. However, the same is not true regarding longer peptides containing the SxIP motif (~11-30 residues) with binding affinities in the low μM range. It is postulated, based on the structural information and the flexibility of the C-terminus tail of EB1cΔ8 that the key to achieve a stronger interaction may be to target an extended region beyond the IP binding site. Therefore, in Chapter 4, efforts were made in order to design a ligand that can interact beyond the IP motif binding site. As an initial approach, a hybrid molecule, composed by a small molecule scaffold followed by a peptide moiety was designed. The chemical scaffold was based on the previously identified IP mimetic and consequently known to be highly specific for EB1's hydrophobic cavity. The peptide scaffold was based on a native ligand's sequence, MACF2, and was composed by the three residues following the SxIP motif – threonine, proline and glutamine. The approach did not yield the expected results, and the specificity for EB1's binding site was lost. A possible cause for this problem is the fact that the IP mimetic scaffold utilised was different from the one initially identified. It included a six-membered ring instead of a five-membered ring and a racemic mixture instead of the previously identified enantiomer. The reason to use a different scaffold from the one previously identified was due to the use of a commercially available scaffold to facilitate and speed up the hybrid molecule synthesis. Despite molecular docking studies demonstrated that the six membered ring would still fit in the binding pocket, it is believed that the loss of specificity to EB1c's binding site was caused due to the larger size of the ring. Another possibility, is that the use of a proline in the peptide sequence attached to the small molecule scaffold could cause the displacement of any other group of the pocket, since the proline residue of the SxIP motif is the only absolutely conserved residue, and replacement of this residue will result in loss of affinity of the SxIP protein to EB1.<sup>17</sup>> Because the proline residue is flanked by two residues with polar side chains – threonine and glutamine, it is very unlikely any of these will go in the hydrophobic binding site of EB1c and therefore reducing the affinity and specificity for it. Finally, the fact that the tetrazole ring is in an axial position to the cyclopentyl group in the

original scaffold, but when using a cyclohexyl ring a bulky substituent will be positioned in equatorial position due to steric hindrance. This seems to be the main reason why the cyclohexyl group is not suitable to replace the cyclopentyl ring. The hybrid molecule approach is an interesting approach where a molecule incorporates a molecular scaffold composed by a fragment previously identified coupled with a peptide derived from a native ligand. Although the approach failed for the reasons outlined above, it would still be of great interest to synthesise another hybrid molecule using the exact fragment identified as the IP mimetic.

Within the same chapter, the process to identify alternatives to the hybrid molecule was described. The methodology followed resembles the one used in Chapter 2, pharmacophore search, docking and multi-parameter analysis, with the main differences being the initial pharmacophore model was built based on the hybrid molecule and the use of *rule of four*<sup>53</sup> instead of *rule of five*<sup>83</sup> (please see description in Chapter 1, section 1.2.1.3) as guidelines for desirable molecular properties. Two additional ligands were identified using this method, but they posed limitations in terms of biophysical methods screening as these are larger molecules with poor aqueous solubility. Lipophilicity is well known to be the contrast of aqueous solubility and lack of solubility has been a consistent problem for medicinal chemists, when it comes to bioavailability.<sup>85</sup> Another reason to avoid large and lipophilic compounds has been demonstrated by Leeson and Springthorpe and postulates that compounds with logP higher than four have increased probability of being promiscuous.<sup>17</sup>

Chapter 5 described in detail the interaction of short length peptide based on a native sequence of a SxIP protein. Contributions to the binding were calculated and detailed in the form of  $\Delta G$  and  $K_d$  values. The SKIP corresponding to the SxIP motif contributes for ~37% of the binding. The extension of this short peptide by two residues, SKIPTP, yields a higher affinity interaction, part given by some extra interactions with the EB1c $\Delta$ 8 C-terminus. The contribution of these two residues is approximately 14%. An eleven residue peptide encompassing the SxIP motif, KPSKIPTPQRK, improves largely the binding affinity from mM range to low  $\mu$ M range. These extra residues KP-----QRK account for the remaining 49% of the interaction. Removing of the C-terminus has a large effect in the interaction of this peptide with EB1c, 19% of the affinity is lost. Based on the important role of the C-terminus a series of three peptides was designed based on the native sequence with

replacements in three amino acids immediately after the SKIP sequence. The mutations involved the replacement of these three amino acids – threonine-proline-glutamine – by hydrophobic amino acids, including leucine, valine and isoleucine. The rationale was to increase the hydrophobic interactions and hence the number of contacts with the C-terminus improving the binding. After docking studies, two of these peptides were selected for testing, respectively KPSKIPLLLRK and KPSKIPVLLRK. Both showed an improved binding affinity with  $K_d$  values of 300 nM and 150 nM respectively. The main difference between these two peptide seems to be that the first leucine after the SKIP sequence seems to have a too bulky side chain, forcing the peptide to adopt a conformation where the backbone of this residue is not accessible to establish hydrogen bond contacts with the backbone of Val254 of the C-terminus, as the wild-type threonine does. The replacement of the leucine to a  $\beta$ -branched side chain residue such as valine seemed to yield the expected results as the enthalpic contributions were improved suggesting the gain of an extra hydrogen bond contribution. Additionally, and to prove the higher affinity of these peptides was achieved by enhancing the contributions with the C-terminus, the first mutant KPSKIPLLLRK was tested against the construct of EB1c lacking the 16 terminal residues – C-terminal tail, EB1c $\Delta$ 16. Whereas for the wild type the difference in the binding with and without the C-terminus was from  $-7.42 \text{ kcal.mol}^{-1}$  to  $6 \text{ kcal.mol}^{-1}$ , the difference is now even larger, from  $-8.91 \text{ kcal.mol}^{-1}$  to  $6.6 \text{ kcal.mol}^{-1}$  (26% instead of 19%), indicating the C-terminus does indeed contribute for the higher affinity interaction observed for this mutant.

Alternatively to this method, solution NMR was the method of choice for screening of all ligands in this project. Additionally, when possible, ITC measurements were also performed as they give complementary information that cannot be measured using NMR. NMR has become a valuable screening tool for the binding of ligands to protein targets, and has the key advantages of being able to detect and quantify interactions with high sensitivity without requiring prior knowledge of protein function. Furthermore, NMR can provide structural information on both the target and the ligand to aid subsequent optimization of weak-binding hits into high-affinity leads.<sup>17</sup>

Important lessons were learned during the course of this project. A single high resolution structure of a protein may not provide all information needed for structure based design, and therefore an ensemble of structures should ideally be used. This is



especially true in proteins with dynamic regions. The detailed understanding of the native interaction in the beginning of the project would very likely yield higher affinity ligands than the ones identified, proving that the acquisition of structural biology data is fundamental for any protein-protein interaction targeting project and should not be dissociated from the drug discovery efforts, but used in an iterative way. This was a mostly iterative project where information obtained from a set of experiments was invariably included in the virtual screening methods. The structural information obtained from EB1c in its free and two different bound forms was fundamental for the understanding of how this protein interacts with SxIP proteins.

In terms of future work that can and should be done, *in vivo* biological assays of the 11 residue peptides, wild-type and the two mutants would be certainly the next step. If the two mutants, with higher binding affinity measured using biophysical techniques can prove their value *in vivo*, disrupting the binding of SxIP protein to EB1 and therefore MT polymerisation, they would give an extremely valuable contribution in understanding the complex protein-protein interaction at the MT plus-ends. Structure elucidation of one or two of the complexes formed between EB1c $\Delta$ 8 and the two mutants would also contribute for a better understanding of how EB1c $\Delta$ 8 interacts with higher affinity ligands. These two peptides have higher affinities to EB1c than any of the natural ligands reported so far. Therefore, upon structural elucidation by solution NMR or X-ray crystallography, a high affinity pharmacophore model could be built and used for the design and identification of higher affinity small molecule modulators for the EB1-SxIP proteins interaction.



## Chapter 7. MATERIALS AND METHODS

### 7.1 Molecular Modelling

The crystal structure of a complex formed between the C-terminal of EB1 lacking the last eight C-terminal residues (EB1c $\Delta$ 8) and a 30 residue peptide derived from the C-terminal of human MACF2 (MACFp1)<sup>31</sup>, with code 3GJO, was downloaded from the RSCB Protein Data Bank (PDB)<sup>102</sup>.

#### 7.1.1 Pharmacophore search

Load the ligand and protein, if required, into ZincPharmer (<http://zincpharmer.csb.pitt.edu>)<sup>55</sup>, using *Load Features* option. Selected options included definition of *Max RMSD* value, *molecular weight* and *number of rotatable bonds*. The used parameters are defined in results section for each performed search.

#### 7.1.2 EB1 Docking Protocol

The crystal structure of the complex (3GJO) is composed of two homodimers, each one with two binding sites and two ligands. For the molecular docking studies just one of the homodimers was used. Additionally, crystal structures of free EB1 (PDB codes 1WU9 and 1YIG) and the ensemble of solution NMR structures obtained during the duration of this project were used for the ensemble docking study described in Chapter 3.

- Protein-ligand molecular docking at the EB1 binding site performed using GOLD 5.0.1<sup>102,111–117</sup>
- Wizard utilised to setup and performing docking calculations.
- Load appropriate file of EB1 protein (pdb format).
- Hydrogen atoms added to the protein using the protonation rules file provided with GOLD
- When specified HOH19 water molecule was extracted for inclusion in docking calculation. All other crystallographic water molecules removed.

- Hydrophobic and hydrogen bonds constraints were used or not depending on the docking protocol.
- MACFp1 ligand loaded and used to define the binding site, together with all atoms around the ligand within 6 Å.
- Ligand file(s) loaded containing the compound(s) to be docked (sdf format). As standard, for each ligand 10 GA runs are performed.
- Select the required fitness scoring function *i.e.* Goldscore. If rescoring is required then select an additional scoring method *e.g.* Chemscore.
- Option for early termination turned off.
- Search efficiency set to 200%. When screening a large number of compounds (*e.g.* PPI-NET library) the *virtual screening* option was used to speed the calculations. This option sets search efficiency to 30%.
- All parameters left as standard, unless otherwise stated.
- Submit calculation and review results

### 7.1.3 Methods used for screening and selection of target molecules

#### 7.1.3.1 Calculation of Molecular properties

Using Pipeline Pilot Professional Client 8.5<sup>16</sup> a range of molecular properties were calculated using the following components *ALogP*, *logD*, *Solubility*, *Surface Area and Volume*, *Molecular Weight*, *Num H Acceptor Donors* and *Molecular Property Counts*. Both input and output files utilised sdf format.

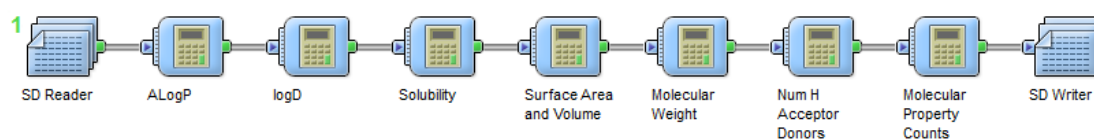


Figure 7.1 – Protocol created in Pipeline Pilot to calculate diverse molecular properties.

### 7.1.3.2 Ligand Efficiency, calculation and ranking

Using KNIME 2.6.3 and the *Math Formula* node with the formula “*score*”/*number of atoms*, ligand efficiency was calculated for each scoring function. The results were ranked in an ascendant order and a new column with a ligand efficiency ranking was added to the SD file.

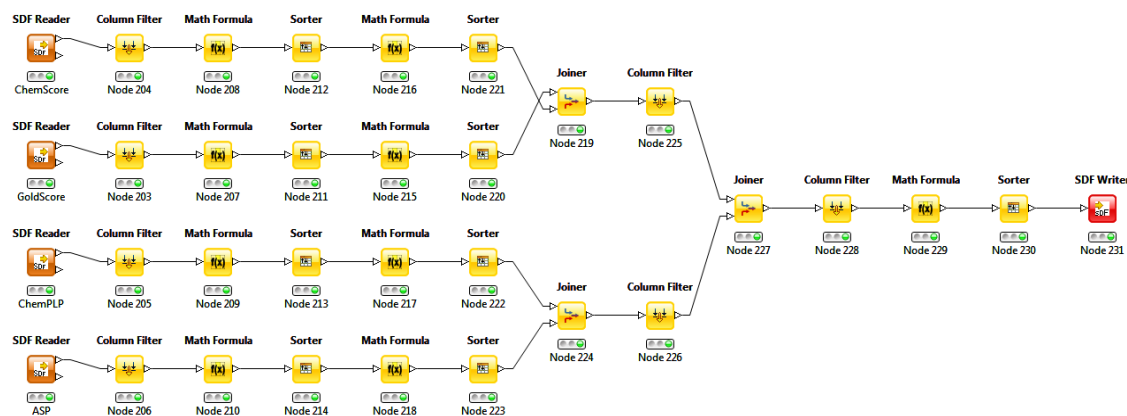


Figure 7.2 – Protocol created using Knime to calculate ligand efficiency for the score results obtained for each scoring function present in GOLD.

### 7.1.3.3 Calculation of docking descriptors

For the calculation of docking descriptors GoldMine was used. GoldMine is a tool embedded in GOLD for the analysis of large quantities of docking information. For that purpose a GoldMine Database was created and the docking results loaded. The following docking descriptors were calculated - *Simple Descriptors*, *General contact descriptors*, *H-bond descriptors* and *Ligand surface area descriptors*. Results were exported in sdf format.

### 7.1.3.4 Multi-objective analysis

Knime 2.6.3 was used for multi-objective analysis, by using the node *Pareto Ranking* to rank the different parameters required for the analysis.



Figure 7.3 – Example of a multi-objective analysis protocol created using Knime and the use of *Pareto Ranking* node.

### 7.1.3.5 RMSD calculation

Align molecules using *substructure* component from Pipeline Pilot was used for RMSD calculation between different docking poses.

### 7.1.3.6 Application of SmartsFilter

Upload a sdf format file into <http://pasilla.health.unm.edu/tomcat/biocomp/smartsfilter> and select the following options – *Blake*, *Glaxo*<sup>16></sup>, *ALARM NMR*<sup>164</sup> and *PAINS*<sup>166</sup>, include pass/fails in the output. Output as sdf format.

### 7.1.4 QSAR experimental procedures

The following procedure was used to develop a QSAR model using Random Forest (RF) or Support Vector Machine (SVM).

- A dataset of 18 molecules tested was re-docked using GOLD<sup>102,111–117</sup>
- Different types of descriptors were calculated for this set of molecules
  - Molecular Fingerprints – FCFP\_2, FCFP\_4, FCFP\_6, ECFP\_2, ECFP\_4, ECFP\_6, Estate Keys and MDLPublicKeys were calculated using Pipeline Pilot's<sup>16></sup> components.
  - Physicochemical descriptors as *ALogP*, *Solubility*, *Surface Area and Volume*, *Molecular weight*, *Num H Acceptor Donors* and *Molecular Property Counts* were calculated in Pipeline Pilot as described previously in the Calculation of Molecular properties section.
  - Derived from docking: these can be divided in two sub-groups: based on the solution files given by GOLD and other descriptors calculated using GoldMine as described previously in the Calculation of docking descriptors section. These are shown in Appendix (A.4).
  - A set of other descriptors were calculated using Dragon 6 – in total 3763 2D descriptors from which 330 were used and 4885 3D descriptors from which 327 were used.
- Calculated descriptors on the set of known active/inactive ligands were used as the training dataset for a machine learning method by using the Pipeline Pilot

components *Learn R SVM Model* and *Learn R Forest Model* for Support Vector Machine and Random Forest models respectively. ROC scores were also calculated for cross-validation of the models for each set of descriptors.

- A sdf format file containing the docking results for the set of compounds to be analysed by these models was uploaded into Pipeline Pilot, filtered using *Model Applicability Filter*, and then screened through each model created previously.

## 7.2 Recombinant EB1 - cloning, expression and purification

### 7.2.1 Materials

#### 7.2.1.1 Water

The methods described in this chapter required at least de-ionised, distilled quality water; this was provided by the Institute of Integrative Biology, University of Liverpool and is termed as RO water in this chapter. For the molecular biology methods detailed in this chapter, ultra-pure quality water was required and this was obtained by use of a Synergy Water Purification System (Millipore) fitted with a SynergyPak® cartridge and a MilliPak-20 Express system; this water is termed as MilliQ in the methods discussed subsequently.

#### 7.2.1.2 General Solvents

Ethanol: Fisher Scientific

Methanol: Fisher Scientific

Isopropanol: Fisher Scientific

Dimethylformamide: Sigma

#### 7.2.1.3 General Reagents

Unless otherwise stated, all reagents used in the methods detailed in this were of laboratory grade and supplied by Sigma-Aldrich.

#### 7.2.1.4 Cell lines

Rosetta(DE3)pLacI, BL21 star (DE3) BL21 (DE3) pLysS, Stellar

#### 7.2.1.5 Antibiotic solutions

Table 7.1 - antibiotic solution composition and concentrations used in this project.

Antibiotic	Stock solution	Solvent	Storage temperature	Working concentration
Kanamycin	32 mg/mL	RO water	-20 °C	32 µg/mL
Chloramphenicol	34 mg/mL	ethanol	-20 °C	34 µg/mL
Ampicillin	50 mg/mL	RO water	-20 °C	50 µg/mL

#### 7.2.1.6 Proteases

The Small ubiquitin-like modifier (SUMO) protease was used to cleave the hexahistidine Ni<sup>2+</sup> affinity tag in the pOPINS (OPPF-UK) vector.

#### 7.2.1.7 Media recipes/composition

The recipes for the different growth media utilised in the methods in this chapter are:

LB Agar

- 37 g of LB agar “Miller” (Merck) per 1 L of RO water

Super Optimal Broth with Catabolite Repression (SOC) medium

- Tryptone (20 g, Fluka)
- Yeast extract (5 g, Fluka)
- NaCl (10 mM, Fisher Scientific)
- KCl (250 mM, ProLabo) , autoclave and then add the following, after filtered sterilised,
- MgSO<sub>4</sub> (10 mM, BDH)
- Glucose (20 mM, BDH)

Lysogeny Broth (LB) medium

- Tryptone
- Yeast extract
- NaCl



2YT medium

- 16 g Tryptone
- 10 g Yeast extract
- 5 g NaCl

Minimal medium - Solution A

- 14.6 g  $\text{Na}_2\text{HPO}_4$
- 5.4 g of  $\text{KHPO}_4$
- 1 g of  $^{15}\text{NH}_4\text{Cl}_2$ , then autoclave and solution B is added after filtered sterilised

Minimal medium - Solution B

- 0.1 M  $\text{MgSO}_4$
- 7.5 mM  $\text{CaCl}_2$
- 0.75 mM  $\text{MnCl}_2$
- 0.25 mM  $\text{FeSO}_4$  (prepared extemporaneously)
- 4 g of Glucose/2 g of  $^{13}\text{C}$  Glucose

**7.2.1.8 Buffers and other solutions**

TAE buffer (50x)

- Tris base (2 M, Fisher Scientific)
- Glacial acetic acid (5.71% (v/v))
- EDTA (5 mM) at pH 8.0

CCMB80 buffer, pH 6.4

- 10 mM KOAc
- 80 mM  $\text{CaCl}_2$
- 20 mM  $\text{MgCl}_2$
- 10% Glycerol
- Filter sterilise and keep at  $\sim 4^\circ\text{C}$

Ni<sup>2+</sup> Affinity Chromatography binding buffer (Low imidazole buffer)

- 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4
- 0.5 M NaCl
- 25 mM Imidazole

Ni<sup>2+</sup> Affinity Chromatography elution buffer (High Imidazole buffer)

- 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4
- 0.5 M NaCl
- 0.5 M Imidazole

Anion exchange - buffer A

- 20 mM Tris, pH 8
- 2 mM DTT

Anion exchange - buffer B

- 20 mM Tris, pH 8
- 2 mM DTT
- 150 mM NaCl

SDS-PAGE gel - loading buffer

- 50 mM Tris/HCl, pH 6.8
- 10% glycerol (v/v)
- 2% SDS (w/v)
- 100 mM DTT
- 0.1% bromophenol blue

SDS-PAGE gel - running buffer (10x)

- Glycine (1.92 M, Fisher Scientific)
- Tris base (250 mM)
- SDS (1% w/v)

#### Coomassie G250 Stain Solution

- Coomassie blue G250 (0.1% (w/v))
- methanol (45% (v/v))
- RO water (45% (v/v))
- acetic acid (10% (v/v))

#### De-stain Solution

- methanol (45% (v/v))
- RO water (45% (v/v))
- acetic acid (10% (v/v))

#### **7.2.1.9 Columns**

HisTrap HP, 5mL (GE Healthcare Life Sciences, 17-2548)

HiTrap Q Sepharose FF (GE Healthcare Life Sciences, 17-5053)

#### **7.2.2 Methods**

The methods presented in this section refer to the methods and techniques used to obtain recombinant EB1 to be used in this project.

##### **7.2.2.1 EB1's EBH Domain Construct Sequences**

The amino acid fragments of EB1 produced for the research described in this thesis are detailed in this sub-section and comprise the EBH domain (C-terminal domain), differing on the inclusion or not of the C-terminal flexible tail. The first construct, named as EB1cΔ8 for practical purposes, includes the sequence between Asp191 and Gly260, whereas the second construct, named as EB1cΔ16 comprises the sequence Asp191-Gly252 - Table 7.2 and Figure 7.4.

Table 7.2 – Parameters calculated using the ProtParam<sup>14</sup> tool and the sequences presented in Figure 7.4.

	EB1cΔ8	EB1cΔ16
Number of amino acids	70	62
Molecular weight (g.mol <sup>-1</sup> )	8043.0	7228.1
Theoretical <i>pI</i>	4.11	4.21
Extinction coefficient ( $\epsilon$ )	2980	2980

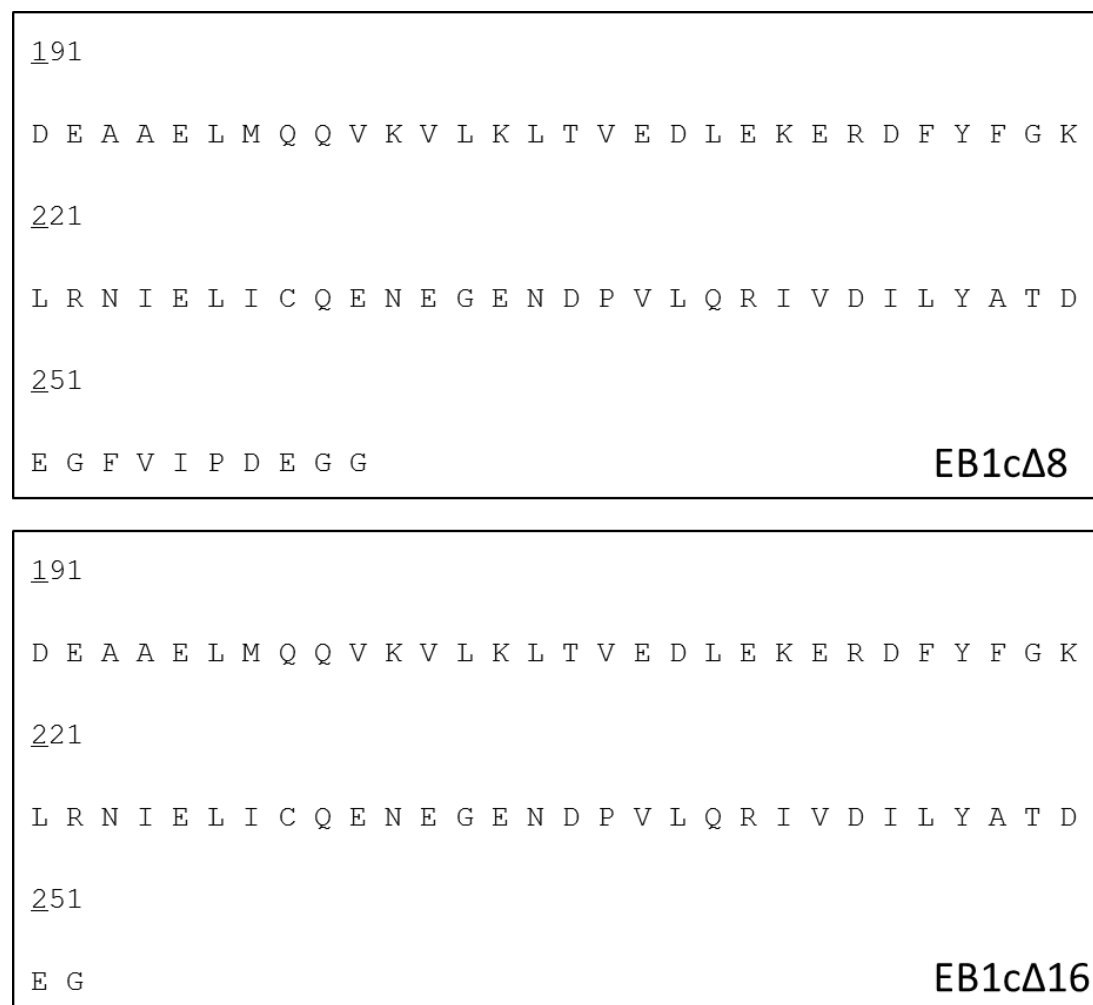


Figure 7.4 – Sequences of the two constructs prepared for EB1c and used in this project.

### 7.2.2.2 Molecular Biology

All PCR fragments were integrated into the OPPF-UK vector suite; all vectors share the same enzyme restriction sites and were linearized using *KpnI* (NEB, #R31425) and *HindIII* (NEB, #R31045) restriction enzymes. EB1cΔ8 fragment cloned into pOPINS and pOPINF was provided by Dr Thomas Zacharchenko and EB1cΔ16 construct was cloned into pOPINS

### 7.2.2.2.1 Vectors used

Table 7.3 - Details of the pOPIN vectors used in this project

Vector	Parent Vector	Antibiotic resistance	Forward Primer Extension	Reverse Primer Extension
pOPINF	pTriEx2	Ampicillin	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINS	pET28a	Kanamycin	GCGAACAGATCGGTGGT	ATGGTCTAGAAAGCTTTA

### 7.2.2.2.2 Primer Design

Primer extensions (Table 7.4) were added in the 3' direction to facilitate integration in the same open reading frame as the fusion tag. Primers were designed to enable synchronous primer annealing to the template DNA with minimal CG content to prevent secondary structure DNA formation. The primers were required to have melting temperatures within 5°C of each other. Additionally, primers were required to be non-complementary to prevent primer dimerization.

Table 7.4 - Primers designed to amplify EB1cΔ16

Name	Sequence (5'3')	Melting temperature (°C)	GC content (%)
EB1Δ16-Fwd	GCGAACAGATCGGTGGTGATGAAGCAGCTGAATTGATGCAGCA	72	51
EB1Δ16-Rev	ATGGTCTAGAAAGCTTTAGCCTTCATCTGTGGCATAAAGAAT	65	38

### 7.2.2.2.3 Polymerase Chain Reaction (PCR)

A PCR reaction was performed for EB1 [191-252] by preparing a 30 μM solution of the primers in nuclease free water and making a “master mix” on ice according with the following table.

Table 7.5 - “Master mix” components and respective volumes used for the PCR reactions performed in this project.

Component	Volume (μL)
10x KOD Hot Start Buffer	5
DNTP mix (2 mM)	5
MgSO <sub>4</sub> (25 mM)	3
Primers (30 μM)	2
DNA template	0.5
KOD Hot Start (1 U/μL)	1
Mili Q water	31.5
Total	50

Using a thermal cycler use the following methodologies

Table 7.6 – temperatures and durations for each step in the PCR reaction.

Step	Temperature	Time
1. Polymerase activation	98°C	2 minutes
2. Denaturation	95°C	30 seconds
3. Annealing	47°C	30 seconds
4. Extension	7 °C	1 minute
<i>30 cycles for steps 2-4</i>		
5. Final Extension	72°C	10 minutes
Hold	4°C	

Hot start polymerases are bound by an antibody and inactive, the initiation of the PCR reaction required pre-incubation for 98°C for 2 minutes, then 95°C for 2 minutes, 47°C - annealing temperature (this varied usually 3°C below the lowest melting temperature of the annealing primer) and a final 72°C extension cycle. The cycle was then repeated 29 times exponentially multiplying the desired DNA fragment. PCR products samples were visualised using 2% (w/v) agarose gel.

The method used was based on the method suggested by NEB (<https://www.neb.com/protocols/2012/10/04/pcr-using-hot-start-taq-dna-polymerase-m0495>).

#### 7.2.2.2.4 Plasmid preparation

The vector pOPINS was transformed into Stellar<sup>TM</sup> Competent Cells in a kanamycin-selective agar plate.

A single colony from the transformed plate was then used to inoculate 10 mL of kanamycin-selective LB and incubated at 37°C for 14-18 hours, with shaking at 180 rpm. The 10 mL culture was then used to inoculate 500 mL of kanamycin-selective LB and incubated at 37°C for nine hours, with shaking at 180 rpm. 500 µL of chloramphenicol (34 mg/mL) were added and the cultures were left to incubate for more 16 hours. When low-copy-number plasmids containing the pMB1 or ColE1 origin of replication are prepared, the yield can be improved by adding chloramphenicol to amplify the copy number. The culture was pelleted by centrifugation and the pellets frozen at -80°C.

The plasmid was extracted and purified using the Plasmid Maxi Kit (QIAGEN, cat no 12162) to yield 100 uL of the plasmid, at the concentrations of 390 ng/µL.

To linearize the vector HindIII-HF and KpnI-HF restriction enzymes were used according with the standard procedure described for these enzymes. The linearized fragment (~5576 kbp) was isolated from a 0.8% Agarose and purified using a gel extraction kit (QIAGEN, cat no 28704).

#### 7.2.2.2.5 Agarose Gel Electrophoresis

For the preparation of the gels for agarose gel electrophoresis (0.8% – 2% (w/v)) 100 mL of 1x TAE buffer and 0.8 g – 2 g of powdered agarose (Bioline) were mixed gradually and heated with periodic swirling to ensure the agarose dissolved. The solution was allowed to cool to < 35°C and the cast for the subsequent gel was created by sealing the ends of the agarose gel plate; 10 µL of ethidium bromide (final concentration = 0.1µg/mL) was then added to the solution. The mixture was poured into the cast, with a sample well comb inserted and allowed to set. Samples to be analysed were mixed in 5:1 ratio with 5x loading buffer and the full volume was pipetted into individual wells. 1 kb standard DNA marker (New England Biolabs) was loaded in a separate well and the gel was run at 60V-90V using a BioRad PowerPac 3000 for ~40 minutes. The gel was then viewed under trans UV light.

#### 7.2.2.2.6 In-Fusion Cloning

For fast and efficient cloning of the PCR products into pOPINS an In-fusion® HD Cloning Kit (Clontech, cat n° 011614) was used. Since the PCR products were not purified a treatment with Cloning Enhancer was performed prior to the In-Fusion Cloning reaction. The only difference to the standard protocol was the volume of the reaction, where a total of 5 µL was used instead of 10 µL – Table 7.7

Table 7.7 – components and respective volumes used for the In-fusion reactions

Component	Volume ( µL)
5X In-Fusion HD Enzyme Premix	1
Linearized vector	1
Purified PCR fragment	1
dH <sub>2</sub> O	2
Total	5

### 7.2.2.3 Protein expression

#### 7.2.2.3.1 Competent Cells

50 µL of commercial glycerol stocks of the BL21 Star<sup>TM</sup> (DE3) and Stellar<sup>TM</sup> *E. coli* competent cells, were used to inoculate 10 mL of LB medium containing kanamycin

(for BL21 Star<sup>TM</sup> (DE3) cells) no antibiotic (for Stellar<sup>TM</sup> cells) and incubated at 37°C for 14-18 hours, with shaking at 180 rpm. 5 mL were used to inoculate 250 mL of LB medium that was subsequently incubated at 37°C with shaking at 180 rpm until reach an optical density at 600 nm (OD<sub>600</sub>) of ~0.4. The culture was then centrifuged at 5000 rpm for 10 minutes at 4°C and the resultant pellet was gently resuspended in 80 mL of CCMB80 buffer. This was then incubated at 0°C for 30 minutes before being centrifuged at 5000 rpm for 10 minutes at 4°C and resuspended in 10 mL of CCMB80. The cells were placed on ice, aliquoted into 50 µL quantities in pre-chilled eppendorfs and stored at -80°C for up to three years.

#### **7.2.2.3.2 Transformation**

~0.4 µL of the desired plasmid was added to 50 µL of competent cells, mixed well and incubated at 0°C for 30 minutes. The cells were then incubated at 42°C for 30 seconds before being incubated at 0°C for 5 minutes. 450 µL of SOC medium was then added and then incubated at 37°C for 1 hour with shaking at 180 rpm. The culture was then centrifuged at 400 rpm for 20 seconds and most of the supernatant was removed, resuspending the pellet in about 50 µL of the remaining supernatant. The transformed cells were then pipetted onto an agar plate (with the necessary antibiotics added, X-Gal or glucose when needed), and spread evenly using an aseptic technique. The plates were then left to incubate at 37°C for ~14-18 hours, without shaking. The agar plates can be stored at 4°C for up to two weeks.

#### **7.2.2.3.3 Expression in 2YT medium**

A single colony from an agar plate was used to inoculate 10 mL of LB medium (with the right antibiotics added) and incubate it for 14-18 hours at 37°C with shaking at 180 rpm. Using the 10 mL of starting culture inoculate 1 L of 2YT medium (with the right antibiotics added) and incubate it at 37°C with shaking at 180 rpm until the OD<sub>600</sub> reaches 0.7-0.8. Cool down the cultures for 30 minutes at 18°C and add IPTG (final concentration 1 mM), leaving cultures at 18°C for 14-18 hours, with shaking at 180 rpm.

The cells were then pelleted by centrifugation, 8000 rpm for 10 minutes at 4°C, discarding the supernatant and resuspending in low imidazole buffer and stored at -80°C until purification.

#### **7.2.2.3.4 Expression in Minimal Medium**

A single colony from an agar plate was used to inoculate 1 mL of kanamycin-selective



LB and incubated at 37°C with shaking at 180 rpm, for 7 hours. 150 µL of the resultant cells were added to 20 mL of kanamycin-selective M9 medium and incubated for 14-18 hours at 37°C, with shaking at 180 rpm. This cell suspension was then added to the 1 L of kanamycin-selective M9 medium and incubated at 37°C until the OD<sub>600</sub> measured between 0.8 – 1. Cool down the cultures for 30 minutes at 18°C and add IPTG (final concentration 0.5 mM), leaving cultures at 18°C for 14-18 hours, with shaking at 180 rpm.

The cells were then pelleted by centrifugation, 8000 rpm for 10 minutes at 4°C, discarding the supernatant and resuspending in low imidazole buffer and stored at -80°C until purification.

#### **7.2.2.4 Protein Purification**

##### **7.2.2.4.1 Cell Lysis**

The cells were gently thawed at room-temperature (20-25°C), and cocktail VII protease inhibitors EDTA free (Calbiochem), bovine deoxyribonuclease (Sigma) and 2-mercaptoethanol (final concentration 1 mM) were added. Cells were mechanically lysed using a Stansted 'Pressure Cell' Homogeniser (SFP Ltd), at 1000 PSI. Cells were then centrifuged at 18000 rpm for 30 minutes and the supernatant subsequently filtered using a 0.22 µm filter.

##### **7.2.2.4.2 Ni<sup>2+</sup> Affinity Chromatography**

A 5 mL HisTrap column (GE Healthcare) was initially washed with MilliQ and then equilibrated with low and high imidazole buffers, at 4 mL/min on an ÄKTApurifier10 FPLC system (GE Healthcare). The filtered cell supernatant was then loaded onto the column at a flow rate of 4 mL/min. The bound contents of the column were then eluted with a gradient of increasing imidazole concentrations at 4 mL/min and fractionated into 5 mL fractions. The fractions were analysed by UV absorbance at 280 nm (A<sub>280</sub>) and those fractions that showed absorption levels above the baseline were analysed using SDS polyacrylamide gel electrophoresis (SDS-PAGE). The fractions that yielded the desired gel retardation band were combined.

##### **7.2.2.4.3 SUMO tag cleavage**

The combined fractions yielded from Ni<sup>2+</sup> Affinity Chromatography were buffer exchanged into low imidazole buffer using a HiPrep<sup>TM</sup> Desalting column. The column was initially washed with MilliQ water and equilibrated using low imidazole buffer.

15 mL of sample were loaded onto the column and the protein was recovered with the same buffer by observing the UV absorbance at 280 nm, yielding a sample with a dilution of 1.2-3 fold. SUMO protease and dithiothreitol (final concentrations of 0.025 µg/mL and 1 mM, respectively) were added to the eluted sample and left at room temperature (20-25°C) for 4 hours. After that time the sample was placed at 4°C for 14-18 hours.

#### **7.2.2.4.4 Reverse Purification**

The sample was loaded into a 5 mL HisTrap column (GE Healthcare) at a 4 mL/min flow rate and readily collected by observing the UV absorbance at 280 nm. The SUMO tag was then eluted by using an increasing concentration of imidazole.

#### **7.2.2.4.5 Ion exchange chromatography**

The sample was exchanged into 20 mM Tris pH 8, 2 mM dithiothreitol, using a HiPrep™ Desalting column and a procedure similar to the one described previously. The sample was then loaded into a 5 mL HiTrap Q (GE Healthcare) column at a flow rate of 4ml/min. The bound contents of the column were then eluted with a linear gradient of 20 mM Tris pH 8, 2 mM dithiothreitol, 1 M NaCl at 4 mL/min and fractionated into 5 mL fractions. The fractions were analysed by UV absorbance at 280 nm ( $A_{280}$ ) and those fractions that showed absorption levels above the baseline were analysed using SDS polyacrylamide gel electrophoresis (SDS-PAGE). The fractions that yielded the desired gel retardation band were combined.

#### **7.2.2.4.6 Concentration of Protein Sample**

Concentration of the combined protein fractions from ion exchange chromatography was carried out using Amicon Ultra Centrifugal Filter units with a 3 kDa molecular weight cut-off and volume capacity of 15 mL (Millipore). The concentration unit membrane was washed by centrifugation (5000 g, swinging bucket rotor) with MilliQ and equilibrated with 20 mM Tris pH 8, 2 mM Dithiothreitol. The protein was then applied to the unit and centrifuged at 5000 g until the desired final volume of sample was achieved. The solution that passed through the membrane (flow through) was analysed by UV absorbance at 280 nm to ensure that there was no membrane failure and hence, that none of the protein of interest had passed through. After use, the concentration units were centrifuged at 5000 g with MilliQ, then 2% (w/v)  $\text{NaN}_3$  and left at 4°C in that solution.

#### **7.2.2.4.7 Buffer Exchange of Protein Sample**

The concentrated fractions from the Ion exchange chromatography were then buffer exchanged into 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 0.02% (w/v)  $\text{NaN}_3$  using a Sephadex G-25 Medium Gravity-Flow PD-10 Column (GE Healthcare). The PD-10 column was initially washed with 25 mL of MilliQ and then 25 mL of 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v)  $\text{NaN}_3$  to equilibrate the column. A maximum of 2.5 mL of the protein sample was applied to the column - for sample volumes less than 2.5 mL, add equilibration buffer to adjust the volume up to 2.5 mL after the sample has entered the packed bed completely. The flow through was discarded and the sample was eluted by addition of 3.5 mL of 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v)  $\text{NaN}_3$ . The final protein solution was flash frozen with liquid nitrogen and stored at  $-80^\circ\text{C}$ .

#### **7.2.2.5 Protein concentration determination**

Protein concentration was determined by absorbance at 280 nm using a nano-drop 3000 and calculated extinction coefficient of respective proteins determined by Prot-Param<sup>14</sup>.

#### **7.2.2.6 SUMO-Protease production**

Recombinant SUMO protease was produced by using a plasmid provided by Dr Paul Elliott in Rosetta pLacI in a ampicilin/chloramphenicol-selective agar plate. One single colony was used to inoculate 20 mL of LB medium and incubated at  $37^\circ\text{C}$  for 14-18 hours. 1 L of 2YT medium (ampicilin/chloramphenicol-selective) was inoculated using the starting culture and incubated until the  $\text{OD}_{600}$  reached 0.6 and then induced using IPTG (final concentration of 500  $\mu\text{M}$ ) and incubated at  $18^\circ\text{C}$  for 14-18 hours. Cells were pelleted by centrifugation and re-suspended in low imidazole buffer containing 10% glycerol, and later lysed as described previously. After nickel affinity chromatography protein was exchanged into 50 mM Tris pH 8, 150 mM NaCl, 3 mM DTT, 5 mM EDTA, 20% glycerol and flash frozen and stored at  $-80^\circ\text{C}$ . Activity of the protease was tested with and verified using SDS-PAGE.

### 7.3 Organic synthesis methods

The materials and methods described in this section relate with the synthesis of molecules tested in this project – compounds **1b** and **2a**.

All reactions were carried out in dry conditions under a nitrogen atmosphere unless otherwise stated.

#### 7.3.1 Materials

Elemental analysis was performed by the microanalysis service at the University of Liverpool. Mass spectra were collected using Micromass LCT Mass Spectrometer or Agilent QTOF 7200 by the mass spectrometry laboratory at the University of Liverpool. <sup>1</sup>H-NMR spectra were recorded on a Bruker AMX 400 (400 MHz) spectrometer, as were <sup>13</sup>C-NMR spectra in solutions of CDCl<sub>3</sub> and MeOD. The chemical shifts are in parts per million (ppm), with tetramethylsilane as the internal reference and the coupling constants in hertz (Hz). TLC was performed on silica plates, and columns were run on silica gel specifically for flash chromatography. Reagents were purchased from SigmaAldrich and Merck.

#### 7.3.2 Methods

##### 7.3.2.1 Compound 1b

Compound **1b** was obtained through a three-step synthesis, as described in Figure 7.5.

Each reaction will be described in the next subsections.

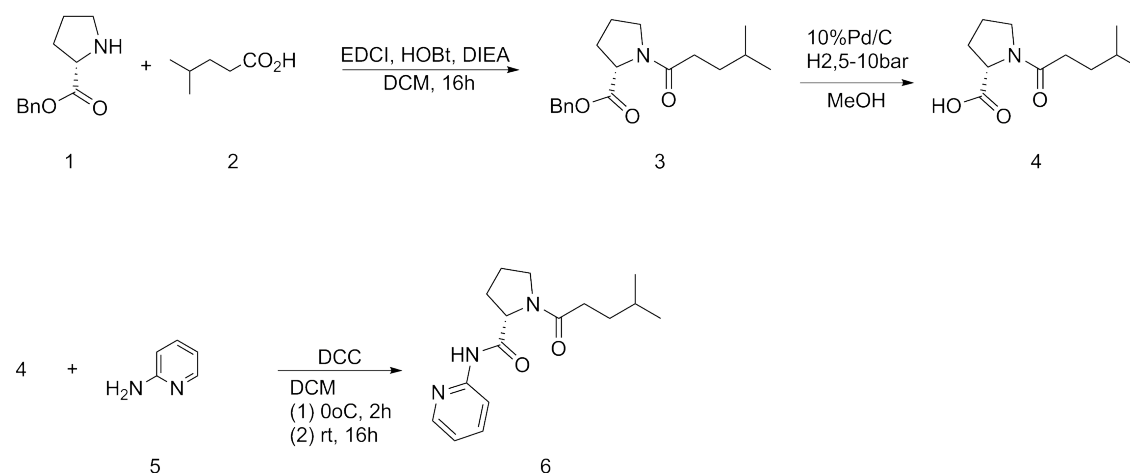


Figure 7.5 – general scheme for the synthesis of compound **1b**.

### 7.3.2.1.1 Synthesis of 3

Proline benzyl ester (1) (556 mg, 2.3 mmol), 4-methyl valeric acid (327  $\mu$ L, 2.6 mmol, 1.1 eq.) and HOBt (398 mg, 2.6 mmol, 1.1 eq.) were dissolved in  $\sim$ 30 mL of anhydrous DCM. After ten minutes EDCI (500 mg, 2.6 mmol, 1.1 eq.) were added, followed by DIEA ten minutes later (881  $\mu$ L, 5.06 mmol, 2.20 eq.). The reaction was allowed to stir overnight. The DCM was removed and the residue was dissolved in EtOAc. The organic layer was washed three times with 1N HCl, three times with saturated NaHCO<sub>3</sub> and three times with brine. The organic layer was dried with Na<sub>2</sub>SO<sub>4</sub>, filtered through filter paper, and concentrated. The product was purified by column chromatography. Fractions were analysed by TLC, concentrated and characterized. 411 mg of product were obtained as a colourless oil (59% yield) <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) $\delta$ : 7.35 (br d, 5H), 5.14 (d, 2H, J=5.7 Hz), 4.47 (m, 1H), 3.63 (m, 2H) 2.36 (m, 2H), 1.96 (m, 4H), 1.58 (m, 1H), 1.47 (m, 2H) 0.91 (dd, 6H, J=6.6, 0.98Hz) *m/z* (LCMS, CI): found 304.19 (M+H)<sup>+</sup>, C<sub>18</sub>H<sub>25</sub>NO<sub>3</sub>, requires 303.18.

### 7.3.2.1.2 Synthesis of 4

The ester 3 was dissolved in MeOH in a Parr bottle. The bottle was flushed with argon and 0.050 eq. of 10% Pd/C. The Parr bottle was placed on a Parr hydrogenation apparatus and subjected to three charge/purge cycles with H<sub>2</sub>. The reaction was then charged with 5-10 bar hydrogen and shaken. After four and a half hours no starting material was observed. The product was dried and 266 mg were obtained as white crystals (92% yield). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) $\delta$ : 4.42 (m, 1H), 3.62 (m, 2H), 2.36 (m, 2H), 2.24 (m, 2H), 2.02 (m, 2H), 1.61 (m, 2H), 1.52 (m, 2H), 0.93 (d, 6H, J=6.52 Hz) CHN analysis: C 59.90%, H 8.58% and N 6.62%. C<sub>11</sub>H<sub>19</sub>NO<sub>3</sub> *m/z* (LCMS, CI): found 214.14 (M+H)<sup>+</sup>, requires 213.14

### 7.3.2.1.3 Synthesis of 6

Compound 4 (266 mg) and 2-aminopyridine (5) (118 mg, 1.25 mmol, 1 eq.) were dissolved in  $\sim$ 7 mL of DCM. The solution was cooled to 0°C, and then DCC (516 mg, 2.5 mmol, 2 eq.) was added. The solution was stirred at 0°C for two hours and at room temperature for another 16 hours. Then the solution was placed in the refrigerator ( $\sim$ 4°C) for two hours, and the white solid was filtered. After removal of solvent under reduced pressure, the residue was purified by column chromatography. 102 mg of a white solid were obtained (29% yield).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ) $\delta$ : 9.68 (br s, 1H), 8.29 (d, 1H, 4.0Hz), 8.15 (d, 1H, 8.3 Hz), 7.66 (td, 1H,  $J=7.7$ ,  $J=1.80$  Hz), 7.00 (dt, 1H,  $J=5.3$ ,  $J=1.80$  Hz), 4.79 (dd, 1H,  $J=8.19$ ,  $J=1.86$  Hz), 3.63 (m, 1H), 3.50 (m, 1H), 2.50 (m, 1H), 2.36 (t, 2H,  $J=7.5$ Hz) 2.06 (m, 4H), 1.60 (m, 2H), 0.91 (d, 6H,  $J=5.92$  Hz)

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ) $\delta$ : 175.70, 171.31, 149.27, 139.58, 120.97, 115.45, 62.19, 49.00, 34.80, 34.07, 29.26, 28.66, 26.47, 23.81, 23.73

CHN analysis C 66.02%, H 8.01%, N 13.62%,  $\text{C}_{16}\text{H}_{23}\text{N}_3\text{O}_2$   $m/z$  (LCMS, CI): found 290.18 ( $\text{M}+\text{H}$ ) $^+$ , requires 289.18

Total yield - 16%

### 7.3.2.2 Compound 2a

#### 7.3.2.2.1 Synthesis of the tri-peptide 2c

All glassware/solvents used for this procedure were thoroughly dried. The dry 2-chlorotryl resin (resin loading 1.51 mmol/g, 0.755mmol) was swelled in 2.5 mL of dry DCM for 15 min. Fmoc-Gln (7) (445 mg, 1.6 eq.) was dissolved in 2 mL of dry DCM with ~0.5 mL of DMF (for solubility) and added to the resin. DIPEA (842  $\mu\text{L}$ , 4 eq. to the amino acid) was added and stirred for two hours. The solvents were then filtered and 0.2 mL of MeOH were added to endcap any remaining trityl groups and stirred for 15 minutes. The resin was then washed twice with DMF, DCM and MeOH (three times). A sample was taken at this stage and checked using the Kaiser test.

The deprotection of the Fmoc group was made by addition of 2 mL of 20% Piperidine/DMF (v/v), followed by a gentle stir for three minutes, removal of the solvent and then a second addition of 20% Piperidine/DMF (v/v), stirred for seven minutes. The solvent was removed by filtration again and 2 mL of DMF were added and left to stir for 30 seconds. This last step was carried out four times. A Kaiser test was performed again at this stage.

The first coupling step was made by addition of Fmoc-Pro (10) (1.55g, 6 eq.) in 6 mL DMF with DIC (600  $\mu\text{L}$ , 5 eq.) and HOBt (518 mg, 5eq.) and left overnight, with gentle agitation. Another deprotection step was performed, using the same method described before.

Fmoc-Thr (13) (1.57 mg, 6 eq.) were added to 6mL of DMF with DIC (600  $\mu\text{L}$ , 5 eq.) and HOBt (518 mg, 5 eq.), and left with agitation for three hours. Another removal of

Fmoc group was performed at this stage.

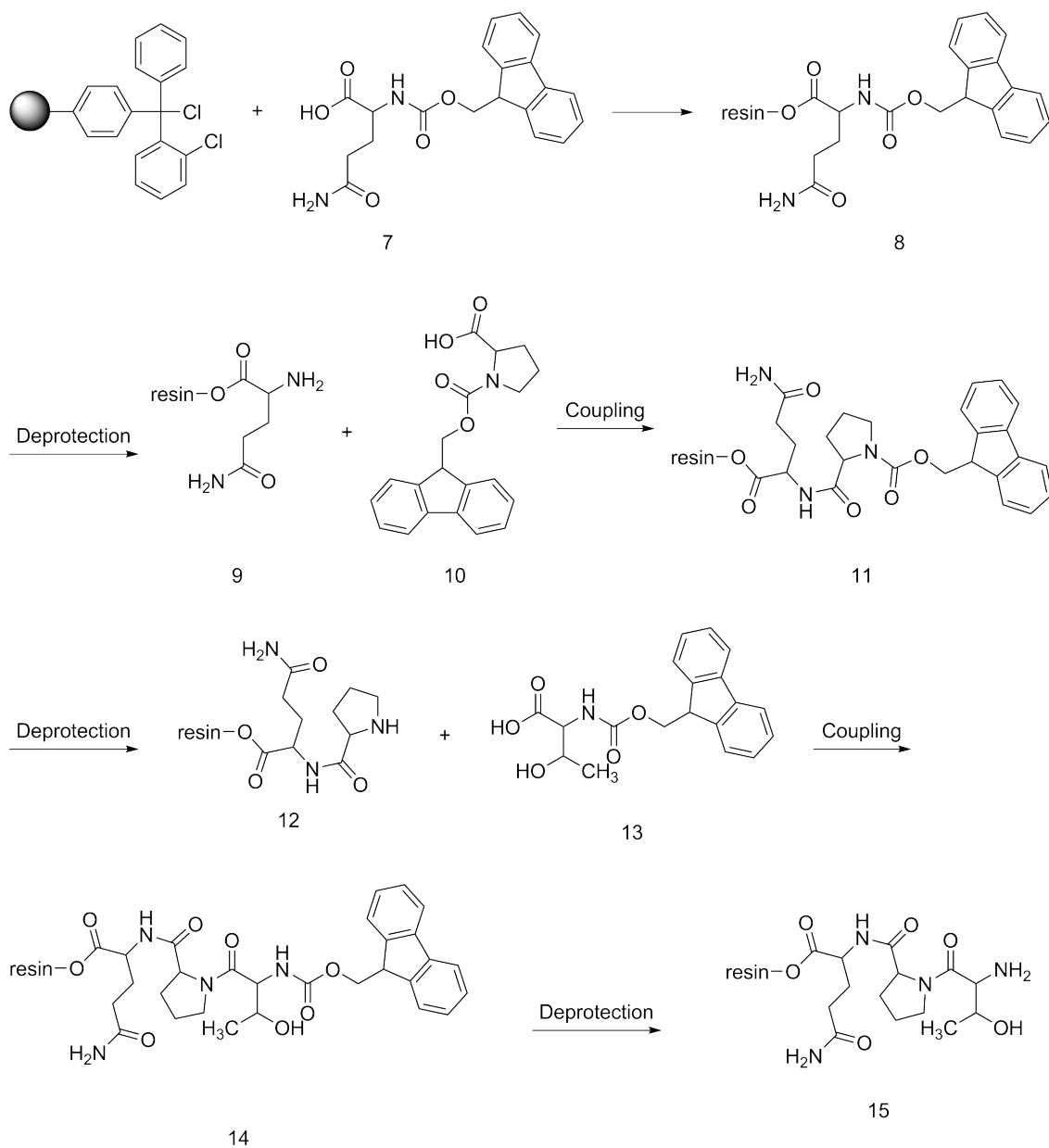


Figure 7.6 – general scheme of the method employed for the synthesis of the tri-peptide TPQ, molecule 2c (15), using solid phase peptide synthesis.





## 7.4 Protein-ligand interactions

This section refers to the materials and methods involved in screening and binding characterisation of the interaction of different ligands and EB1. The ligands tested include peptides and small molecules. The stock solutions prepared for these ligands were used for both NMR and ITC.

Synthetic peptides derived from the C-terminus of the MACF2 (GL Biochem and ChinaPeptides, China) were directly resuspended into the identical buffer as the protein - 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v) NaN<sub>3</sub> - and minimal pH adjustments were made if required.

100-250 mM stock solutions of the potential EB1 binder were prepared by dissolving the appropriate mass of the organic compound in the necessary amount of hexadeuterodimethyl sulfoxide (DMSO-d<sub>6</sub>).

### 7.4.1 NMR Spectroscopy

#### 7.4.1.1 General

Samples to be examined by NMR spectroscopy of volumes 600 µL were pipetted into Wilmad 5 mm 175 mm 800 MHz Precision Pyrex Glass NMR Tubes (Goss Scientific). Samples of limited volume were pipetted into Shigemi Advanced NMR Microtube Assembly – Matched with <sup>2</sup>H<sub>2</sub>O, bottom L 8 mm tubes (Sigma-Aldrich).

The NMR spectra were acquired at 25°C, unless otherwise stated, on either a Bruker AVANCE II+ 600 MHz Ultrashield or 800 MHz US<sup>2</sup> spectrometer, equipped with triple resonance cryoprobes. The Bruker TopSpin programme version 3.1 was used to process the resultant NMR spectra and the Collaborative Computational Project for NMR Analysis (CcpNmr Analysis)<sup>17</sup> software was used for interactive spectral analysis and assignment.

#### 7.4.1.2 Ligand Binding Screening by NMR Spectroscopy

1D <sup>1</sup>H<sup>179,18</sup> and 2D <sup>1</sup>H-<sup>15</sup>N HSQC<sup>179-181</sup> NMR experiments were performed using 0.05 mM or 0.1 mM of <sup>15</sup>N-labelled EB1 in 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v) NaN<sub>3</sub>. Two different methods were used, one for the first set of tested molecules and another for all the remaining compounds.

For the first set of molecules the required volume of each organic compound stock solution was added to the protein sample to achieve the desired ratios. This method was also used for the longest peptide sequences tested – 11 residues (11MACF, 11MACF-LLL and 11MACF-VLL).

For all the remaining molecules, the method was altered in order to facilitate the comparison between experiments by keeping the volume of DMSO-d6 constant. This method consisted by preparing two samples with the same concentration of DMSO-d6, one without ligand and the other with the maximum ligand concentration, and by mixing them in the right proportions obtain a range of titration points. The maximum DMSO-d6 concentration achieved was 4% (v/v). For the shorter length peptides – 4MACF and 6MACF a similar method was used however without any DMSO added. Because these are weaker binders when compared with 11MACF, the volume to be added to achieve the highest ratio would dilute the sample and therefore affect the results.

The obtained spectra were then assigned by transfer of the previously determined unbound backbone assignments of EB1 and analysed to identify which residue(s) had been perturbed. The chemical shift perturbation value ( $\Delta\delta$  ppm) for each residue was then calculated – Equation 8.1<sup>94,18></sup>,

$$\Delta\delta_{ppm} = \sqrt{((\Delta\delta_{HN})^2 + (\Delta\delta_N * \alpha_N)^2)}, \text{ Equation 8.1}$$

where,  $\alpha_N$  = Scaling Factor of 0.15.

#### 7.4.1.3 Binding Affinity Determination by NMR Spectroscopy

To determine the binding affinity ( $K_d$ ) by NMR spectroscopy the same experiments used to screen the ligands were used – see previous section.

The individual residue chemical shifts for each titration point were assigned and fitted using Equation 8.2<sup>183,18></sup>,

$$y = A((B + x) - \sqrt{((B + x)^2 - 4x)}), \text{ Equation 8.2}$$

where,  $x = \frac{[L]}{[P]}$ ,  $y = \Delta\delta$ ,  $A = \frac{1}{2}\Delta\delta_{max}$  and  $B = 1 + \left(\frac{K_D}{[P]}\right)$

### 7.4.2 Isothermal Titration Calorimetry (ITC)

All isothermal titration calorimetry experiments were carried out at 25°C on an iTC200 Microcalorimeter (GE Healthcare) with a 200 µL cell capacity and 40 µL syringe volume. For all experiments protein, either EB1cΔ8 or EB1cΔ16 in 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v) NaN<sub>3</sub> were in the cell, and the ligand was in the syringe. The ITC titration data collected was analysed using the Origin®7 software.

The lyophilised peptides were directly resuspended into the same buffer as the protein and minimal pH adjustments were made if required.

Table 7.8 - Conditions for ITC experiments performed with EB1cΔ8

ligand	ligand concentration (µM)	protein concentration (µM)	injection volumes (µL)	number of injections
11MACF	750	50	1.5	25
11MACF_LLL	225	15	1.5	25
11MACF_VLL	250	25	1.5	25

Table 7.9 - Conditions for ITC experiments performed with EB1cΔ16

ligand	ligand concentration (µM)	protein concentration (µM)	injection volumes (µL)	number of injections
11MACF	2250	150	1.5	25
11MACF_LLL	2250	150	1.5	25

### 7.4.3 Octet® RED96

The compounds screened by this method belong to a library of compounds – PPI-Net library, and each compound was at a concentration of 10 mM in DMSO. One hundred thirty-five compounds were screened at a final concentration of 0.5 mM – Figure 7.8.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A2	A3	A4	A5	A6	A7	A8
B																	
C	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B2	B3	B4	B5	B6	B7	B8
D																	
E	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C2	C3	C4	C5	C6	C7	C8
F																	
G	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D2	D3	D4	D5	D6	D7	D8
H																	
I	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E2	E3	E4	E5	E6	E7	E8
J																	
K	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F2	F3	F4	F5	F6	F7	F8
L																	
M	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G2	G3	G4	G5	G6	G7	G8
N																	
O	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H2	H3	H4	H5	H6	H7	H8
P																	




 - 20 mM PO<sub>4</sub>, 50 mM NaCl, 0.5 mM TCEP, 0.02% NaN<sub>3</sub> + 0.01% TWEEN +5% DMSO  
 - Plate 20000604 + 20 mM PO<sub>4</sub>, 50 mM NaCl, 0.5 mM TCEP, 0.02% NaN<sub>3</sub> + 0.01% TWEEN  
 - Plate 20000605 + 20 mM PO<sub>4</sub>, 50 mM NaCl, 0.5 mM TCEP, 0.02% NaN<sub>3</sub> + 0.01% TWEEN

Figure 7.8 – schematic representation of the plate prepared for Octet® RED96 screening.

#### 7.4.4 NMR based structure calculation

For this section the materials and equipment are the same as described in section 7.4.1.1.

##### 7.4.4.1 Resonance assignment for backbone and side-chain

3D HNCO, HN(CA)CO, HNCA, HNCACB and CBCACONH experiments were used for the sequential assignment of the backbone NH, N, CO, C $\alpha$  and C $\beta$  resonances. Side chain assignments were obtained using 3D HBHA(CO)NH, H(C)CH-TOCSY and (H)CCH-TOCSY experiments. Aromatic side-chains were assigned using 2D-NOESY and <sup>13</sup>C-resolved 3D NOESY-HSQC. The resonances of the ligands were assigned using <sup>13</sup>C,<sup>15</sup>N-filtered 2D TOCSY and NOESY experiments.

The HNCACB<sup>136,139,186</sup> experiment correlates each NH group in the protein with its own <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts and the respective <sup>13</sup>C $\alpha/\beta$  chemical shifts of the preceding residue (*i-1* and *i*); the CBCACONH<sup>139,186</sup> correlates a particular NH group in the protein with the <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts of the preceding residue (*i-1*) (Figure 7.9). The correlation observed in this complementary set of spectra allowed

the sequential linking of one NH group to the next (Figure 7.9, top panel). The assignment of the  $^{13}\text{C}$  (carbonyl) was accomplished using a similar approach.  $\text{HNCO}^{138,140,187,188}$  shows the correlation between NH and the carbonyl for the preceding residue and  $\text{HN}(\text{CA})\text{CO}^{141,188}$  correlates the NH with the respective residue and the preceding ( $i-1$  and  $i$ ) - Figure 7.10. By superimposing the two later spectra we can link the NH resonances of the residues. The  $\text{HNCA}^{138,140,189}$  shows for each NH strip the  $\text{C}\alpha$  for the respective residue and for the preceding one ( $i-1$  and  $i$ ) and it is useful to complement the assignment, this experiment was only used for backbone assignment of the EB1-11MACF complex.

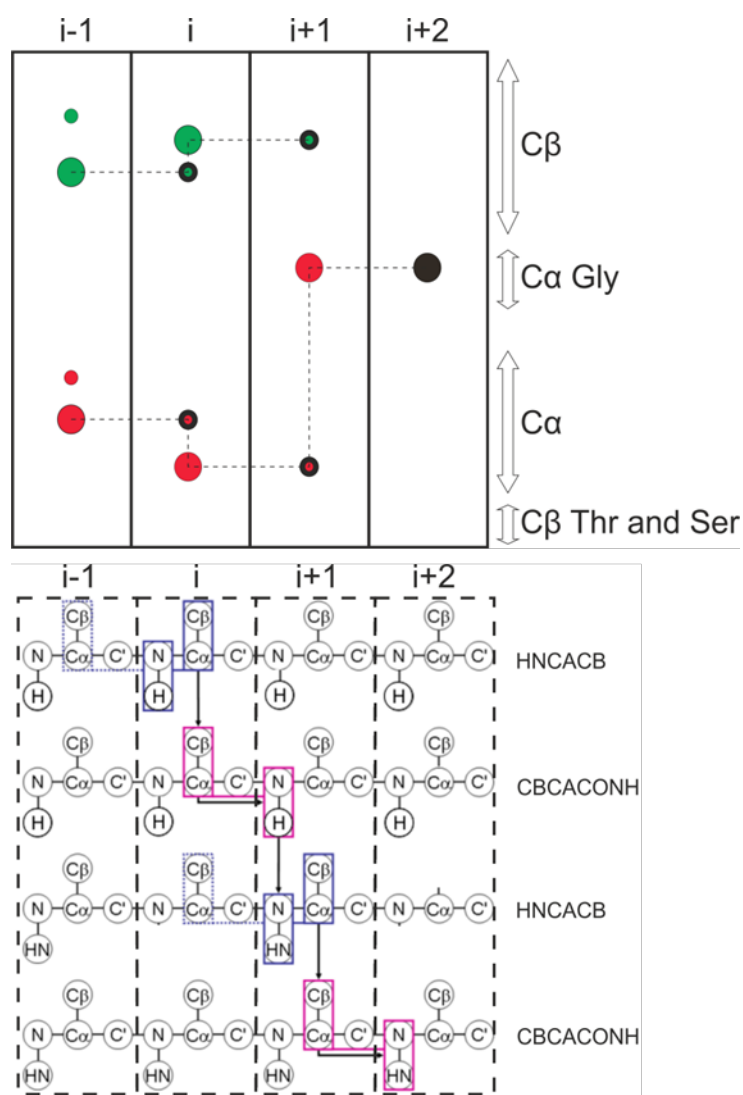


Figure 7.9 - Top panel – Schematic representation of the pair CBCA(CO)NH/HNCACB. CBCA(CO)NH spectrum is shown in black, with the  $\text{C}\alpha/\text{C}\beta$  peaks of the HNCACB spectrum are shown in green and red respectively. The bottom panel refers to the resonances that can be observed when using this pair of experiments and how are they linked to perform sequential assignment.

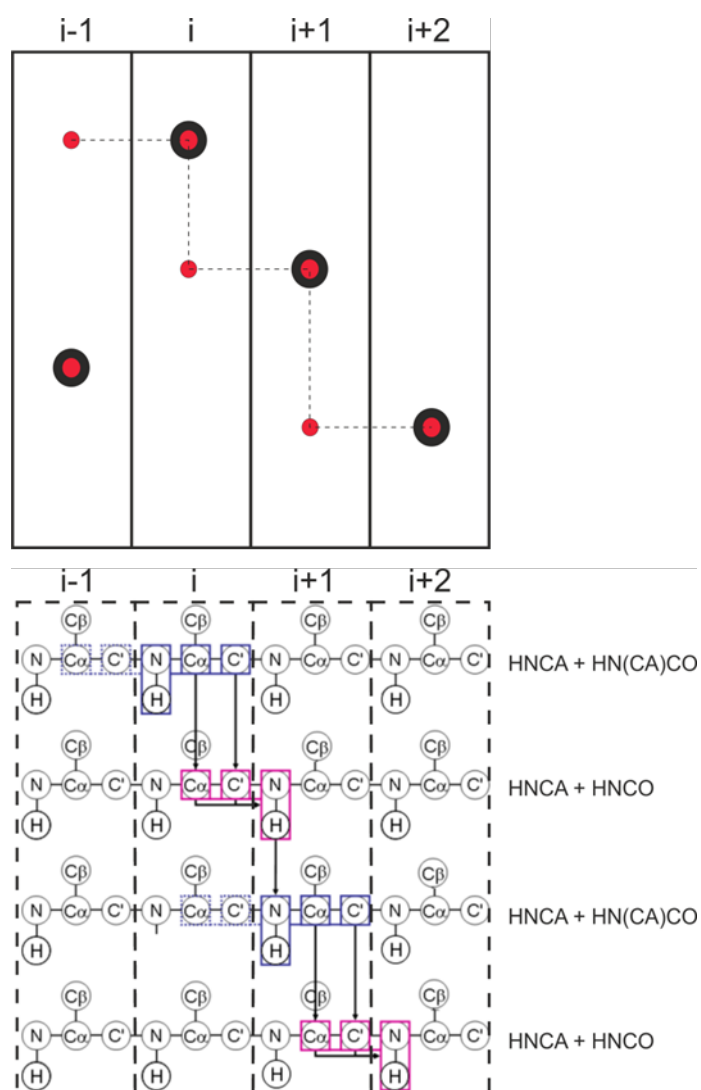


Figure 7.10 - Top panel – Schematic representation of the pair HN(CA)CO/HNCO used for CO assignment performed by sequential assignment; the HN(CA)CO spectrum is shown in red, with the CO peak of the preceding residue being less intense and overlapped with the HNCO spectrum showed in black. The bottom panel refers to the resonances that can be observed using this pair of experiments and how are they linked to perform sequential assignment.

#### 7.4.4.2 Structure calculation of free EB1c18

A 1 mM of uniformly isotopically labelled  $^{13}\text{C}$ ,  $^{15}\text{N}$ -EB1 was prepared in 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v)  $\text{NaN}_3$ . A large number of experiments were acquired using this sample and are detailed in Table 7.10. Because EB1 is a symmetric homodimer the intermonomer contacts of the coiled coil/leucine zipper are fundamental for structure elucidation and characterisation. Therefore, a sample of  $^{13}\text{C}$ ,  $^{15}\text{N}$ -EB1 and unlabelled EB1 at equimolar concentrations was prepared and incubated for 16 hours at  $37^\circ\text{C}$ .<sup>42</sup> The spectra acquired to characterise the

intermonomer contacts are detailed in Table 7.11.

Table 7.10 – Spectra acquired in a double labelled sample ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) for EB1c $\Delta$ 8 resonance assignment

NMR experiment	No. of Scans	Mixing time (ms)	No. of points digitized			FID resolution (Hz)			Spectrometer frequency
			F1	F2	F3	F1	F2	F3	
1D $^1\text{H}^{178,17}$	64	-	8192	-	-	2.06	-	-	600 MHz
1D $^1\text{H}$ – Excitation Sculpting <sup>18</sup>	16	-	32768	-	-	0.59	-	-	600 MHz
$^1\text{H}$ - $^{15}\text{N}$ -HSQC <sup>179-18</sup> >	4	-	512	1400	-	19.00	9.43	-	600 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aliphatic) <sup>19</sup> >	4	-	180	1200	-	105.63	11.00	-	600 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC-constant time (aliphatic) <sup>19</sup> >	4	-	458	1200	-	41.52	11.00	-	600 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aromatic)	8	-	64	1024	-	94.33	11.74	-	600 MHz
$^1\text{H}$ - $^{13}\text{C}$ -TROSY (aromatic) <sup>19</sup> >	8	-	88	1024	-	68.60	11.74	-	600 MHz
3D-HNCO (nus) <sup>140,187,18</sup> >	2	-	128	76	1200	23.58	43.21	10.02	600 MHz
3D-HNCACO <sup>141,18</sup> >	2	-	128	76	1200	10.00	27.00	10.02	600 MHz
3D-CBCACONH <sup>136,139,18</sup> >	2	-	114	80	1200	57.00	27.00	10.02	600 MHz
3D-HNCACB (nus) <sup>137,139,18</sup> >	4	-	1280	80	1200	143.36	41.05	10.02	600 MHz
3D-HBHACONH <sup>139,18</sup> >	8	-	200	76	1100	33.01	41.61	13.11	600 MHz
H(C)CH-TOCSY <sup>193-19</sup> >	4	-	300	128	1024	23.21	146.19	10.57	600 MHz
(H)CCH-TOCSY	4	-	128	128	1024	146.19	146.19	10.57	600 MHz
$^{15}\text{N}$ -resolved- $^{13}\text{C}$ -decoupled-NOESY-HSQC <sup>196,19</sup> >	2	100	512	160	1400	31.26	26.36	11.45	800 MHz
$^{13}\text{C}$ -resolved-NOESY-HSQC <sup>196,19</sup> >	2	200	512	180	1400	31.26	138.64	11.45	800 MHz
$^{13}\text{C}$ -resolved-NOESY-HSQC (aromatic)	8	100	450	52	1200	35.57	154.82	13.36	800 MHz

Table 7.11 - Spectra acquired in an unlabelled/labelled sample for intermonomer resonance assignment.

NMR experiment	No. of Scans	Mixing time (ms)	No. of points digitized			FID resolution (Hz)			Spectrometer frequency
			F1	F2	F3	F1	F2	F3	
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-NOESY	80	200	512	2048	-	31.26	7.83	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered-2D-NOESY	48	200	400	2048	-	40.01	7.83	-	800 MHz
2D-NOESY*	72	200	512	2048	-	31.26	7.83	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC	8	200	440	68	1400	32.74	165.73	11.45	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC*	8	200	440	64	1400	32.74	176.09	11.45	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC (aromatic)	8	200	400	48	1024	40.02	142.56	12.52	800 MHz

\*spectra acquired in 100% <sup>2</sup>H<sub>2</sub>O (all other experiments acquired with 5-10% <sup>2</sup>H<sub>2</sub>O)

The majority of restraints for NMR structure calculations are provided by NOESY derived distance restraints. However, the NOE assignment is tedious due to the large number of assignment possibilities, peak overlap and potential artefacts in the spectra. Manual assignment is time consuming and may not provide an accurate structure due to human error. In order to hasten the process and reduce the risk of errors, NOE assignment was carried out using the program ARIA 2.3.1 (Ambiguous Restraints for Iterative Assignment)<sup>96</sup>. ARIA uses an iterative protocol and the concept of ambiguous distance restraints (ADR)<sup>19</sup> to automatically assign NOE cross-peaks. It does not perform the structure calculations itself but assigns NOE cross-peaks by first deriving all possible assignments for each peak by matching a list of chemical shifts with frequency ‘windows’ centred around the position of a peak in an iterative structure calculation scheme. Structures are calculated using the program CNS (Crystallography & NMR System)<sup>98,199</sup>, ARIA then analyses the conformers obtained in order to update the restraints and obtain a set of improved conformer.

For structure calculation using ARIA two main methods for generation of distance restraints can be generally used: restraints can be generated by ARIA itself or restraints can be generated in CcpNmr Analysis<sup>17</sup> and imported into ARIA. In this project a combination of both was used.

EB1 cross-peaks and peak lists were imported from CcpNmr Analysis into ARIA, and restraints generated by ARIA for the <sup>15</sup>N-resolved-<sup>13</sup>C-decoupled-NOESY-HSQC,



$^{13}\text{C}$ -resolved-NOESY-HSQC,  $^{15}\text{N}$ ,  $^{13}\text{C}$ -filtered- $^{13}\text{C}$ -resolved-NOESY-HSQC (intermonomer contacts),  $^{13}\text{C}$ -resolved-NOESY-HSQC for aromatic resonances. The tolerances for direct proton, indirect proton and heteronuclear dimensions were defined to 0.02, 0.04 and 0.5, respectively. Spin diffusion correction was also used, by inputting the spectrometer frequency, mixing time and correlation time for each spectrum –Table 7.12.

Table 7.12 – Spectra used in ARIA for NOE iterative assignment and distance restraints calculation.

		$^{15}\text{N}$ -resolved- $^{13}\text{C}$ -decoupled-NOESY-HSQC	$^{13}\text{C}$ -resolved-NOESY-HSQC	$^{15}\text{N}$ , $^{13}\text{C}$ -filtered- $^{13}\text{C}$ -resolved-NOESY-HSQC	$^{13}\text{C}$ -resolved-NOESY-HSQC (aromatics)
Use manual assignments		Yes	Yes	Yes	Yes
Enable structural rules		Yes	No	No	No
Tolerances	Proton 1	0.02			
	Hetero 1	0.5			
	Proton 2	0.04			
	Hetero 2	0.5			
Molecule correlation time (ns)		10			
Spectrometer frequency (MHz)		800			
Mixing time (ms)		100	200	200	100
Ambiguity level (for multimers)		Intra-molecular only	Unknown	Intermolecular only	Unknown

All NOESY peaks were picked semi-automatically in CcpNmr Analysis with noise and artefact peaks removed manually. Cross-peak intensities were used to evaluate the target distances. The ambiguity of these restraints derived from the NOE cross peaks was resolved in the ARIA cycles with the violation tolerances set to 3.0-0.1 Å for iterations 1-8, respectively.

As mentioned before, distance restraints generated in CcpNmr Analysis were also directly used for the structure calculation, namely dihedral angles, restraints hydrogen bond restraints and intermonomer restraints.

Table 7.13 – Distance restraints derived by CcpNmr Analysis and imported to ARIA to perform the structure calculations.

Restraint type	Number of restraints
Dihedral	128
Hydrogen bond	90
Intermonomer – aliphatic region	83
Intermonomer – aromatic region	16

Dihedral restraints ( $\phi/\psi$ ) were generated using DANGLE (Dihedral ANGles from Global Likelihood Estimates)<sup>20</sup> module of CcpNmr Analysis software. DANGLE predicts protein backbone  $\phi$  and  $\psi$  angles and secondary structure assignments solely from amino acid sequence information, experimental chemical shifts and a database of known protein structures and their associated shifts. A total of 128 restraints were obtained using DANGLE.

Hydrogen bond restraints between the N-H group of an amino acid and the C=O group of the amino acid *four* residues earlier ( $i + 4 \rightarrow i$ ) were manually inserted in CcpNmr Analysis for the regions where the secondary structure is defined as alpha-helical – [191-231] and [237-247]. There is a strong reasoning for the use of these “artificial” restraints. First of all the NOE data and dihedral angles data were used to predict the secondary structure in CcpNmr Analysis and the prediction is not only in accordance with previous EB1 published structures<sup>24,27,31,44,48,20</sup>, but also with experimental data acquired. The 90 restraints included the default “N-H..O=C” and their co-linear restraints.

Distance restraints between atoms in chain A and chain B – intermonomer, were generated from assigned peaks from the isotope filtered <sup>13</sup>C-NOESY-HSQC and 2D NOESY. In total, 99 restraints were used.

Finally, 320 structures were calculated, 30 best refined in the presence of explicit water molecules as described in Table 7.14, 20 lowest structure used for the analysis and PDB deposition. For speed, structures are typically calculated “in vacuum” which can result artefacts as a result of the simplified treatment of non-bonded forces and missing solvent contacts. ARIA provides the option of refining structures in a shell of water molecules for a defined number of structures from the final iteration with a full molecular dynamics force field incorporating electrostatics. This refinement helps to avoid unrealistic side chain packing and unsatisfied hydrogen bond donors or receptors.

To ensure there are no systematic differences that could influence validation results, the force fields used for water refinement (file PARALLHDG 5.3) are consistent with the force fields used for structure calculation and validation.<sup>20></sup>

Table 7.14 – Parameters used for each iteration of structure calculation, including number of structures calculated, use or not of spin diffusion, violation tolerance, and ambiguity cut-off.

Iteration	No. of Structures	Keep n best structures	Use best n structures	Spin diffusion	Violation tolerance Å	Partial assignment – ambiguity cut-off
0	64	10	7	No	3.0	1.0
1	64	10	7	No	3.0	0.9999
2	64	10	7	No	3.0	0.999
3	64	10	7	No	1.0	0.99
4	64	10	7	No	1.0	0.98
5	64	10	7	Yes	1.0	0.96
6	64	10	7	Yes	0.5	0.93
7	160	10	7	Yes	0.5	0.9
8	320	20	7	Yes	0.1	0.8
Water refinement	30	-	-	-	-	-

Table 7.15 - Simulating annealing parameters used for structure determination of free EB1cΔ8

Type	Torsion angle
Random seed	89764443
TAD high temperature	10000.0
TAD time-step factor	9.0
Cartesian High temperature	2000.0
Cartesian 1 <sup>st</sup> iteration	0
Time-step	0.00015
Cool1 final temperature	1000.0
Cool2 final temperature	50.0
High-temp steps	20000
Refine steps	8000
Cool1 steps	10000
Cool2 steps	8000

#### 7.4.4.3 Structure calculation of EB1cΔ8 – 1a complex

For this complex it was known, from acquired data – e.g. NH intensities, secondary structure prediction, NOEs – that the structure would not be very different from the free protein.

Table 7.16 - Spectra acquired in a double-labelled sample ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) for EB1cΔ8-1a complex resonance assignment.

NMR experiment	No. of Scans	Mixing time (ms)	No. of points digitized			FID resolution (Hz)			Spectrometer frequency
			F1	F2	F3	F1	F2	F3	
1D $^1\text{H}$	64	-	8192	-	-	2.54	-	-	800 MHz
1D $^1\text{H}$ – Excitation Sculpting	32	-	32768	-	-	0.78	-	-	800 MHz
$^{15}\text{N}$ , $^{13}\text{C}$ -filtered-1D $^1\text{H}^*$	256	-	8192	-	-	2.73	-	-	800 MHz
$^1\text{H}$ - $^{15}\text{N}$ -HSQC	4	-	360	1600	-	11.72	11.00	-	800 MHz
3D-CBCACONH	4	-	156	102	1400	147.06	41.35	11.45	800 MHz
3D-HBACONH	2	-	200	102	1400	44.01	41.35	11.45	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aliphatic)	8	-	200	1400	-	12.58	126.90	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC-constant time (aliphatic)	4	-	612	1400	-	12.58	41.47	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aromatic)	16	-	64	1200	-	13.36	125.79	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -TROSY (aromatic)	16	-	1200	118	-	13.36	68.23	-	800 MHz
H(C)CH-TOCSY	2	-	128	48	1024	23.21	146.19	11.74	800 MHz
HCCH-TOCSY	4	-	128	128	1024	146.19	146.19	11.74	800 MHz
$^{15}\text{N}$ -resolved- $^{13}\text{C}$ -decoupled-TOCSY-HSQC	2	45	400	102	1024	40.02	37.76	15.65	800 MHz
$^{15}\text{N}$ -resolved- $^{13}\text{C}$ -decoupled-NOESY-HSQC	2	150	512	112	1400	31.26	36.21	11.45	800 MHz
$^{13}\text{C}$ -resolved-NOESY-HSQC	2	200	512	180	1200	31.26	138.54	13.36	800 MHz
$^{13}\text{C}$ -resolved-NOESY-HSQC*	4	100	512	80	1400	28.14	140.87	11.45	800 MHz
$^{13}\text{C}$ -resolved-NOESY-HSQC (aromatic)	4	200	450	48	1200	35.57	167.72	13.36	800 MHz
$^{13}\text{C}$ -decoupled-2D-TOCSY*	96	45	512	2048	-	31.26	7.83	-	800 MHz
$^{13}\text{C}$ -decoupled-2D-NOESY*	128	200	600	2048	-	31.25	7.83	-	800 MHz
$^{15}\text{N}$ , $^{13}\text{C}$ -filtered- $^{13}\text{C}$ -decoupled-2D-NOESY	96	200	400	2048	-	40.00	7.83	-	800 MHz
$^{15}\text{N}$ , $^{13}\text{C}$ -filtered- $^{13}\text{C}$ -decoupled-2D-NOESY*	64	200	512	2018	-	31.25	7.83	-	800 MHz

<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC	8	200	380	64	1400	37.92	176.16	11.45	800 MHz
--	---	-----	-----	----	------	-------	--------	-------	---------

\*spectra acquired in 100% <sup>2</sup>H<sub>2</sub>O (all other experiments acquired with 5-10% <sup>2</sup>H<sub>2</sub>O)

Therefore we used the final set of restraints – ambiguous and unambiguous - from the calculation of the structure of the free protein and introduced intermolecular restraints between the small molecule and the protein. Concerning the dihedral angles restraints and hydrogen bond restraints the same principle applies. The key for the calculation of the structure of this complex is to use as best as we can the intermolecular NOEs (protein – small molecule). Topology file for the small molecule **1a** was obtained via CcpNmr ACPYPE Portal<sup>20</sup>.

The <sup>15</sup>N, <sup>13</sup>C-filtered-<sup>13</sup>C-resolved-NOESY-HSQC was carefully assigned and the restraints directly used without any filtering. Regarding the intermolecular NOEs from <sup>15</sup>N, <sup>13</sup>C-filtered-<sup>13</sup>C-decoupled-2D-NOESY experiment, the assigned restraints were also used but it was allowed to ARIA to filter not compatible restraints due to the absence of a carbon dimension and the inherent ambiguity. In addition to these restraints, restraints derived from assigned peaks of <sup>13</sup>C-decoupled-2D-NOESY for the aromatic region were also used.

Table 7.17 – Parameters used for each iteration of structure calculation, including number of structures calculated, use or not of spin diffusion, violation tolerance, and ambiguity cut-off.

Iteration	No. of Structures	Keep n best structures	Use best n structures	Spin diffusion	Violation tolerance Å	Partial assignment – ambiguity cut-off
0	32	0	7	No	3.0	1.0
1	32	0	7	No	3.0	0.9999
2	32	0	7	No	3.0	0.999
3	32	0	7	No	1.0	0.99
4	32	0	7	No	1.0	0.98
5	32	0	7	No	1.0	0.96
6	32	0	7	No	0.5	0.93
7	32	0	7	No	0.5	0.9
8	160	0	20	No	0.1	0.8
Water refinement	20	-	-	-	-	-

For this calculation an initial structure model was also included. This model came from

previous attempt for determine the structure of this complex and had an overall energy of -4703 kcal. For that reason it was necessary to modify some dynamic parameters related with the simulating annealing protocol, as decreasing the temperature from 10000 to 2000, and increasing the cooling steps Table 7.18. It has been shown slower SA cooling protocols improves the quality of the structures obtained and therefore the success of the structure calculation.<sup>15></sup>

Table 7.18 - Simulating annealing parameters used for structure determination of free EB1cΔ8

Type	Torsion angle
Random seed	89764443
TAD high temperature	2000.0
TAD time-step factor	9.0
Cartesian High temperature	2000.0
Cartesian 1 <sup>st</sup> iteration	0
Time-step	0.00075
Cool1 final temperature	1000.0
Cool2 final temperature	50.0
High-temp steps	40000
Refine steps	16000
Cool1 steps	40000
Cool2 steps	32000

#### 7.4.4.4 Structure calculation of EB1cΔ8 - 11MACF complex

To 277 μL of 1.15 mM uniformly <sup>13</sup>C, <sup>15</sup>N-EB1cΔ8 in 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v) NaN<sub>3</sub>, 73 μL of 11MACF stock solution (5.35 mM) were added, making a final ratio of 1:1. To this sample more 11MACF was then added in order to get a 1:2 ratio. A second sample of 1.15 mM uniformly <sup>13</sup>C, <sup>15</sup>N-EB1cΔ8 was prepared in 20 mM phosphate pH 6.5, 50 mM NaCl, 0.5 mM TCEP, 0.02% (w/v) NaN<sub>3</sub> and the following samples were prepared by adding the necessary amount of 5.35 mM 11MACF in a sequential manner:

Table 7.19 – titration of 11MACF into  $^{13}\text{C}$ ,  $^{15}\text{N}$ -EB1cΔ8 and unlabelled EB1cΔ8 in order to obtain the best quality spectra possible for both protein and ligand resonances, *i.e.* the line broadening is minimised.

$^{13}\text{C}$ , $^{15}\text{N}$ -EB1cΔ8:11MACF ratio	EB1cΔ8:11MACF ratio
10:1	10:9
10:2	10:10
10:4	10:11
10:9	
10:10	
10:11	

For each ratio both 1D  $^1\text{H}$  and  $^{15}\text{N}$ ,  $^{13}\text{C}$ -filtered-1D  $^1\text{H}$  were acquired in order to find a ratio where both protein and peptide peaks were sharp enough to acquire NOE data. The 10:11 ratio showed the best compromise in terms of sharp peaks for the peptide and sharp peaks for protein (Appendix A.3).

Table 7.20 - Spectra acquired in a double labelled sample ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) for EB1cΔ8-11MACF complex at different ratios for resonance assignment.

NMR experiment	Ratio	No. of Scans	Mixing time	No. of digitised points			FID resolution (Hz)			Spectrometer frequency
				F1	F2	F3	F1	F2	F3	
1D $^1\text{H}$	1:1	64	-	8192	-	-	2.54	-	-	800 MHz
$^{15}\text{N}$ , $^{13}\text{C}$ -filtered-1D $^1\text{H}$	1:1	64	-	8192	-	-	1.96	-	-	800 MHz
$^1\text{H}$ - $^{15}\text{N}$ -HSQC	1:1	4	-	360	1600	-	11.72	11.00	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aliphatic)	1:1	4	-	200	1400	-	126.90	12.58	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC-constant time (aliphatic)	1:1	4	-	612	1400	-	41.47	12.58	-	800 MHz
$^1\text{H}$ - $^{13}\text{C}$ -HSQC (aromatic)	1:1	8	-	64	1024	-	94.35	11.74	-	600 MHz
$^1\text{H}$ - $^{13}\text{C}$ -TROSY (aromatic)	1:1	8	-	88	1024	-	68.60	11.74	-	600 MHz
3D-HNCO (nus)	1:1	2	-	128	100	1400	31.45	42.17	11.45	800 MHz
3D-HNCACO	1:1	2	-	96	100	1400	41.93	42.17	11.45	800 MHz
3D-CBCACONH (nus)	1:1	2	-	156	102	1400	147.06	41.35	11.45	800 MHz
3D-HNCACB (nus)	1:1	2	-	156	102	1400	147.06	12.58	12.58	800 MHz
3D-HBHACONH	1:1	2	-	230	102	1400	38.27	41.35	11.45	800 MHz
H(C)CH-TOCSY	1:1	2	-	300	128	1024	23.21	146.30	10.57	600 MHz
HCCH-TOCSY	1:1	4	-	128	128	1024	146.30	146.30	10.57	600 MHz
$^{15}\text{N}$ , $^{13}\text{C}$ -filtered- $^{13}\text{C}$ -decoupled-2D-	1:1	80	45	512	2048	-	31.25	7.83	-	800 MHz

TOCSY										
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-NOESY	1:1	96	200	512	2048	-	31.25	7.83	-	800 MHz
1D <sup>1</sup> H	1:2	64	-	8192	-	-	2.54	-	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered-1D <sup>1</sup> H	1:2	64	-	8192	-	-	1.96	-	-	800 MHz
<sup>15</sup> N-resolved-HSQC	1:2	2	-	360	1600	-	11.72	11.00	-	800 MHz
<sup>13</sup> C-resolved-HSQC (aliphatics)	1:2	4	-	200	1400	-	63.06	12.58	-	800 MHz
<sup>13</sup> C-resolved-HSQC-constant time (aliphatics)	1:2	4	-	612	1400	-	63.06	11.00	-	800 MHz
3D-CBCACONH	1:2	8	-	156	102	1400	147.06	41.35	11.45	800 MHz
3D-HNCA <sup>138,140,18&gt;</sup>	1:2	8	-	80	98	1400	138.36	43.03	11.45	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-TOCSY	1:2	80	45	512	2048	-	31.25	7.83	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-NOESY	1:2	96	200	512	2048	-	31.26	7.83	-	800 MHz
<sup>15</sup> N-resolved- <sup>13</sup> C-decoupled-NOESY-HSQC	1:2	2	150	512	152	1400	31.25	27.75	11.45	800 MHz
1D <sup>1</sup> H	10:11	64	-	8192	-	-	2.54	-	-	800 MHz
1D <sup>1</sup> H – Excitation Sculpting*	10:11	128	-	8192	-	-	2.73	-	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered-1D <sup>1</sup> H	10:11	256	-	8192	-	-	2.73	-	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered-1D <sup>1</sup> H*	10:11	256	-	8192	-	-	2.73	-	-	800 MHz
<sup>1</sup> H- <sup>15</sup> N-HSQC	10:11	2	-	360	1600	-	11.72	11.00	-	800 MHz
<sup>1</sup> H- <sup>15</sup> N-HSQC*	10:11	4	-	360	1600	-	11.72	11.00	-	800 MHz
<sup>1</sup> H- <sup>13</sup> C-HSQC (aliphatic)	10:11	8	-	200	1400	-	126.79	12.58	-	800 MHz
<sup>1</sup> H- <sup>13</sup> C-HSQC-constant time (aliphatic)	10:11	4	-	612	1400	-	41.43	12.58	-	800 MHz
<sup>1</sup> H- <sup>13</sup> C-HSQC-constant time (aliphatic)*	10:11	4	-	612	1400	-	41.47	12.58	-	800 MHz
<sup>1</sup> H- <sup>13</sup> C-HSQC (aromatic)	10:11	16	-	64	1200	-	125.80	13.36	-	800 MHz
<sup>1</sup> H- <sup>13</sup> C-TROSY (aromatics)	10:11	16	-	118	1200	-	68.23	13.36	-	800 MHz
<sup>13</sup> C-decoupled-2D-NOESY*	10:11	16	60	512	1440	-	31.25	11.13	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-	10:11	80	45	512	2048	-	31.25	7.83	-	800 MHz



decoupled-2D-TOCSY										
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-TOCSY*	10:11	64	40	512	1440	-	31.25	11.13	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered-2D-NOESY	10:11	48	200	512	2048	-	31.25	7.83	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-NOESY	10:11	48	200	512	2048	-	31.25	7.83	-	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-decoupled-2D-NOESY*	10:11	80	200	512	1440	-	31.25	11.13	-	800 MHz
<sup>13</sup> C-resolved-NOESY-HSQC	10:11	2	200	512	172	1400	31.25	144.99	11.45	800 MHz
<sup>13</sup> C-resolved-NOESY-HSQC*	10:11	2	200	512	80	1400	28.14	140.87	11.45	800 MHz
<sup>13</sup> C-resolved-NOESY-HSQC (aromatics)	10:11	4	200	450	56	1200	35.56	143.78	13.36	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC	10:11	8	200	440	68	1400	32.83	165.79	11.45	800 MHz
<sup>15</sup> N, <sup>13</sup> C-filtered- <sup>13</sup> C-resolved-NOESY-HSQC*	10:11	8	200	440	80	1400	32.75	140.87	11.45	800 MHz

\* spectra acquired in 100% <sup>2</sup>H<sub>2</sub>O (all other experiments acquired with 5-10% <sup>2</sup>H<sub>2</sub>O)

In order to perform the structure calculation of the complex the parameters used for free EB1cΔ8 were used (Table 7.14 and Table 7.15). Three spectra were iteratively assigned by ARIA, <sup>15</sup>N-resolved-<sup>13</sup>C-decoupled-NOESY-HSQC and two <sup>13</sup>C-resolved-NOESY-HSQC experiments with two different mixing times, 200 and 100 ms -

Table 7.21. In terms of restraints introduced into ARIA as CcpNmr Analysis calculated restraints from manual assignments or obtained from previous structure calculations as .tbl files can be found in Table 7.22. These include dihedral and hydrogen bond restraints, similarly to what was described for the previous structures, intermolecular restraints between EB1 and 11MACF assigned from filtered experiments and intermonomer restraints. The latter include the 99 restraints used in the calculations of free EB1cΔ8 resultant from manual assignments but also 249 restraints resultant from the final iteration of the same structure, yielding a total of 348 intermonomer restraints.

Table 7.21 – Spectra used in ARIA for NOE iterative assignment and distance restraints calculation.

		<sup>15</sup> N-resolved- <sup>13</sup> C-decoupled- NOESY-HSQC	<sup>13</sup> C-resolved- NOESY-HSQC	<sup>13</sup> C-resolved- NOESY-HSQC
Use manual assignments		Yes	Yes	Yes
Enable structural rules		Yes	No	No
Tolerances	Proton 1	0.02		
	Hetero 1	0.5		
	Proton 2	0.04		
	Hetero 2	0.5		
Molecule correlation time (ns)		10		
Spectrometer frequency (MHz)		800		
Mixing time (ms)		100	200	100
Ambiguity level (for multimers)		Intra-molecular only	Unknown	Unknown

Table 7.22 – Distance restraints derived from CcpNmr Analysis or ARIA (calculation of the structure of free EB1cΔ8) to perform the structure calculations of the complex.

Restraint type	Number of restraints
Dihedral	130
Hydrogen bond	90
Intermonomer – aliphatic region	83
Intermonomer – aromatic region	16
Intermonomer *	249
Intermolecular	185

\* obtained from ARIA assignment for free EB1cΔ8.

## REFERENCES

1. Huber, F., Boire, A., López, M. P. & Koenderink, G. H. Cytoskeletal crosstalk: when three different personalities team up. *Curr. Opin. Cell Biol.* (2015). doi:10.1016/j.ceb.2014.10.005
2. Akhmanova, A. & Steinmetz, M. O. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nat. Rev. Mol. Cell Biol.* **9**, 309–322 (2008).
3. Gardner, M. K., Zanic, M. & Howard, J. Microtubule catastrophe and rescue. *Curr. Opin. Cell Biol.* **25**, 14–22 (2013).
4. Nogales, E. & Wang, H.-W. W. Structural intermediates in microtubule assembly and disassembly: how and why? *Curr. Opin. Cell Biol.* **18**, 179–184 (2006).
5. Alfaro-Aco, R. & Petry, S. Building the Microtubule Cytoskeleton Piece By Piece. *J. Biol. Chem.* **290**, jbc.R115.638452 (2015).
6. Mitchison, T. & Kirschner, M. Dynamic instability of microtubule growth. *Nature* **312**, 237–242 (1984).
7. Burbank, K. S. & Mitchison, T. J. Microtubule dynamic instability. *Curr. Biol.* **16**, 516–517 (2006).
8. Zhang, R., Alushin, G. M., Brown, A. & Nogales, E. Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. *Cell* **162**, 849–859 (2015).
9. Nogales, E. & Heald, R. Microtubule dynamics. *J. Cell Sci.* **115**, 3–4 (2002).
10. Desai, A. & Mitchison, T. J. Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol.* **13**, 83–117 (1997).
11. Mandelkow, E.-M. E. & Mandelkow, E.-M. E. Microtubule structure. *Curr. Opin. Struct. Biol.* **4**, 171–179 (1994).
12. Galjart, N. Plus-End-Tracking Proteins and Their Interactions at Microtubule Ends. *Curr. Biol.* **20**, R528–R537 (2010).
13. Slep, K. C. & Vale, R. D. Structural Basis of Microtubule Plus End Tracking by XMAP215, CLIP-170, and EB1. *Mol. Cell* **27**, 976–991 (2007).
14. Howard, J. & Hyman, A. A. Dynamics and mechanics of the microtubule plus end. *Nature* **422**, 753–758 (2003).
15. Jiang, K. & Akhmanova, A. Microtubule tip-interacting proteins: a view from both ends. *Curr. Opin. Cell Biol.* **23**, 94–101 (2011).
16. Yu, I., Garnham, C. P. & Roll-Mecak, A. Writing and Reading the Tubulin Code. *J. Biol. Chem.* **290**, jbc.R115.637447 (2015).
17. Gouveia, S. M. & Akhmanova, A. *Cell and molecular biology of microtubule plus end tracking proteins. End binding proteins and their partners. International Review of Cell and Molecular Biology* **285**, (2010).
18. Honnappa, S. *et al.* Key interaction modes of dynamic +TIP networks. *Mol. Cell* **23**, 663–71 (2006).
19. Maurer, S. P., Fourniol, F. J., Bohner, G., Moores, C. A. & Surrey, T. EBs Recognize a Nucleotide-Dependent Structural Cap at Growing Microtubule Ends. *Cell* **149**, 371–382 (2012).

20. Roberts, A. J., Goodman, B. S. & Reck-Peterson, S. L. Reconstitution of dynein transport to the microtubule plus end by kinesin. *Elife* **2014**, (2014).
21. Akhmanova, A. & Steinmetz, M. Microtubule +TIPs at a glance. *J Cell Sci* **123**, 3415–3419 (2010).
22. Kumar, P. & Wittmann, T. +TIPs: SxIPping along microtubule ends. *Trends Cell Biol.* **22**, 418–428 (2012).
23. Kanaba, T. *et al.* Microtubule-binding sites of the CH domain of EB1 and its autoinhibition revealed by NMR. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 499–507 (2013).
24. Bjelić, S. *et al.* Interaction of mammalian end binding proteins with CAP-Gly domains of CLIP-170 and p150(glued). *J. Struct. Biol.* **177**, 160–7 (2012).
25. Abiatari, I. *et al.* The microtubule-associated protein MAPRE2 is involved in perineural invasion of pancreatic cancer cells. *Int. J. Oncol.* **35**, 1111–1116 (2009).
26. Komarova, Y. *et al.* Mammalian end binding proteins control persistent microtubule growth. *J. Cell Biol.* **184**, 691–706 (2009).
27. Slep, K. C. *et al.* Structural determinants for EB1-mediated recruitment of APC and spectraplakins to the microtubule plus end. *J. Cell Biol.* **168**, 587–598 (2005).
28. Yan, S. *et al.* Three-dimensional structure of CAP-gly domain of mammalian dynactin determined by magic angle spinning NMR spectroscopy: conformational plasticity and interactions with end-binding protein EB1. *J. Mol. Biol.* **425**, 4249–66 (2013).
29. Duellberg, C. *et al.* Reconstitution of a hierarchical +TIP interaction network controlling microtubule end tracking of dynein. *Nat. Cell Biol.* **16**, 804–811 (2014).
30. Slep, K. C. Structural and mechanistic insights into microtubule end-binding proteins. *Curr. Opin. Cell Biol.* **22**, 88–95 (2010).
31. Honnappa, S. *et al.* An EB1-Binding Motif Acts as a Microtubule Tip Localization Signal. *Cell* **138**, 366–376 (2009).
32. Kumar, P. *et al.* Multisite Phosphorylation Disrupts Arginine-Glutamate Salt Bridge Networks Required for Binding of Cytoplasmic Linker-associated Protein 2 (CLASP2) to End-binding Protein 1 (EB1). *J. Biol. Chem.* **287**, 17050–64 (2012).
33. Maki, T., Grimaldi, A. D., Fuchigami, S., Kaverina, I. & Hayashi, I. CLASP2 Has Two Distinct TOG Domains That Contribute Differently to Microtubule Dynamics. *J. Mol. Biol.* **427**, 2379–2395 (2015).
34. Al-Bassam, J., Larsen, N. a., Hyman, A. A. & Harrison, S. C. Crystal Structure of a TOG Domain: Conserved Features of XMAP215/Dis1-Family TOG Domains and Implications for Tubulin Binding. *Structure* **15**, 355–362 (2007).
35. Al-Bassam, J. & Chang, F. Regulation of microtubule dynamics by TOG-domain proteins XMAP215/Dis1 and CLASP. *Trends Cell Biol.* **21**, 604–614 (2011).
36. Grimaldi, A. D. *et al.* CLASPs are required for proper microtubule localization of end-binding proteins. *Dev. Cell* **30**, 343–352 (2014).
37. Kim, M.-J. *et al.* Depletion of end-binding protein 1 (EB1) promotes apoptosis of human non-small-cell lung cancer cells via reactive oxygen species and Bax-mediated mitochondrial

- dysfunction. *Cancer Lett.* **339**, 15–24 (2013).
38. Su, L. *et al.* APC Binds to the Novel Protein EB1 Advances in Brief APC Binds to the Novel Protein EB1. 2972–2977 (1995).
  39. Liu, M. *et al.* EB1 acts as an oncogene via activating  $\beta$ -catenin/TCF pathway to promote cellular growth and inhibit apoptosis. *Mol. Carcinog.* **48**, 212–219 (2009).
  40. Jiang, K. *et al.* A Proteome-wide screen for mammalian SxIP motif-containing microtubule plus-end tracking proteins. *Curr. Biol.* **22**, 1800–7 (2012).
  41. Stroud, M. J., Kammerer, R. a. & Ballestrom, C. Characterization of G2L3 (GAS2-like 3), a new microtubule- and actin-binding protein related to spectraplakins. *J. Biol. Chem.* **286**, 24987–24995 (2011).
  42. Buey, R. M. *et al.* Insights into EB1 structure and the role of its C-terminal domain for discriminating microtubule tips from the lattice. *Mol. Biol. Cell* **22**, 2912–2923 (2011).
  43. Honnappa, S., John, C. M., Kostrewa, D., Winkler, F. K. & Steinmetz, M. O. Structural insights into the EB1-APC interaction. *EMBO J.* **24**, 261–9 (2005).
  44. Huang, X., Lovelace, L. ., Smith, D. & Lebioda, L. Structure of fragment of human end-binding protein 1 (EB1) containing the N-terminal domain at 1.35 Å resolution. *To be Publ.* doi:10.2210/pdb2r8u/pdb
  45. Hayashi, I. & Ikura, M. Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1). *J. Biol. Chem.* **278**, 36430–36434 (2003).
  46. Tomizawa, T., Kigawa, T., Koshiba, S., Inoue, M. & Yokoyama, S. Solution structure of the CH domain from mouse EB-1. *To be Publ.* doi:10.2210/pdb1v5k/pdb
  47. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–12 (2014).
  48. Honnappa, S., John, C. M., Kostrewa, D., Winkler, F. K. & Steinmetz, M. O. Structural insights into the EB1-APC interaction. *EMBO J.* **24**, 261–269 (2005).
  49. Buey, R. M. *et al.* Sequence determinants of a microtubule tip localization signal (MtLS). *J. Biol. Chem.* **287**, 28227–42 (2012).
  50. Scolz, M. *et al.* GTSE1 Is a Microtubule Plus-End Tracking Protein That Regulates EB1-Dependent Cell Migration. *PLoS One* **7**, (2012).
  51. Berlow, R. B., Dyson, H. J. & Wright, P. E. Functional advantages of dynamic protein disorder. *FEBS Lett.* **589**, 2433–2440 (2015).
  52. Azzarito, V., Long, K., Murphy, N. S. & Wilson, A. J. Inhibition of  $\alpha$ -helix-mediated protein-protein interactions using designed molecules. *Nat. Chem.* **5**, 161–73 (2013).
  53. Morelli, X., Bourgeas, R. & Roche, P. Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* **15**, 475–481 (2011).
  54. Sperandio, O. O., Miteva, M. A. & Villoutreix, B. O. in *Virtual Screening* (ed. Sotriffer, C.) 435–465 (Wiley-VCH Verlag GmbH & Co. KGaA, 2011). doi:10.1002/9783527633326.ch16
  55. Koes, D. R. & Camacho, C. J. ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res.* **40**, W409–W414 (2012).
  56. Zhong, S., Macias, A. T. & MacKerell Jr., A. D. Computational identification of inhibitors of

- protein-protein interactions. *Curr Top Med Chem* **7**, 63–82 (2007).
57. J. Wilson, A. Inhibition of protein-protein interactions using designed molecules. *Chem. Soc. Rev.* **38**, 3289–3300 (2009).
  58. Sperandio, O., Reynès, C. H., Camproux, A.-C. & Villoutreix, B. O. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today* **15**, 220–9 (2010).
  59. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009 (2007).
  60. Pelay-Gimeno, M., Glas, A., Koch, O. & Grossmann, T. N. Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chemie Int. Ed.* n/a-n/a (2015). doi:10.1002/anie.201412070
  61. Jubb, H., Higuero, A. P., Winter, A. & Blundell, T. L. Structural biology and drug discovery for protein protein interactions. *Trends Pharmacol. Sci.* **33**, 241–248 (2012).
  62. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem. Biol.* **21**, 1102–1114 (2014).
  63. Jin, L., Wang, W. & Fang, G. Targeting protein-protein interaction by small molecules. *Annu. Rev. Pharmacol. Toxicol.* **54**, 435–56 (2014).
  64. Tse, C. *et al.* ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. *Cancer Res.* **68**, 3421–3428 (2008).
  65. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009 (2007).
  66. Eyrisch, S. & Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **50**, 3457–3464 (2007).
  67. Falchi, F., Caporuscio, F. & Recanatini, M. Structure-based design of small-molecule protein – protein interaction modulators : the story so far. *Future Med. Chem.* **6**, 343–357 (2014).
  68. Winter, A. *et al.* Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery. *Q. Rev. Biophys.* **45**, 383–426 (2012).
  69. Skwarczynska, M. & Ottmann, C. Protein-protein interactions as drug targets. *Future Med. Chem.* **7**, 2195–2219 (2015).
  70. Bortolato, A. & Moro, S. Designing a ligand for pharmaceutical purposes. *Expert Opin. Drug Discov.* **3**, 579–590 (2008).
  71. Leach, A. R. & Gillet, V. J. *An Introduction to Chemoinformatics*. (Springer, 2007). doi:10.1017/CBO9781107415324.004
  72. Kolb, P., Ferreira, R. S., Irwin, J. J. & Shoichet, B. K. Docking and chemoinformatic screens for new ligands and targets. *Curr. Opin. Biotechnol.* **20**, 429–36 (2009).
  73. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
  74. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and

- recent advances. *Drug Discov. Today* **15**, 444–450 (2010).
75. Sanders, M. P. A. *et al.* Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **52**, 1607–1620 (2012).
  76. Zoete, V., Grosdidier, A. & Michielin, O. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell. Mol. Med.* **13**, 238–248 (2009).
  77. Gowthaman, R., Deeds, E. J. & Karanicolas, J. Structural properties of non-traditional drug targets present new challenges for virtual screening. *J. Chem. Inf. Model.* **53**, 2073–81 (2013).
  78. Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **303**, 1813–1818 (2004).
  79. Moro, S., Bacilieri, M. & Deflorian, F. Combining ligand-based and structure-based drug design in the virtual screening arena. *Expert Opin. Drug Discov.* **2**, 37–49 (2007).
  80. Oprea, T. I. & Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **8**, 349–358 (2004).
  81. Clark, D. E. What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discov.* **3**, 841–851 (2008).
  82. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142 (2014).
  83. Lipinski, C. a, Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
  84. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
  85. Hann, M. M. Molecular obesity, potency and other addictions in drug discovery. *Multifaceted Roles Crystallogr. Mod. Drug Discov.* **2**, 183–196 (2015).
  86. Schultes, S. *et al.* Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today. Technol.* **7**, e147-202 (2010).
  87. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev Drug Discov.* **3**, 1474–1776 (2004).
  88. Grünberg, R., Nilges, M. & Leckner, J. Flexibility and Conformational Entropy in Protein-Protein Binding. *Structure* **14**, 683–693 (2006).
  89. Grünberg, R., Leckner, J. & Nilges, M. Complementarity of structure ensembles in protein-protein binding. *Structure* **12**, 2125–2136 (2004).
  90. Gul, S. & Hadian, K. Protein-protein interaction modulator drug discovery: past efforts and future opportunities using a rich source of low- and high-throughput screening assays. *Expert Opin. Drug Discov.* **9**, 1393–404 (2014).
  91. Silvestre, H. L., Blundell, T. L., Abell, C. & Ciulli, A. Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12984–9 (2013).
  92. Carlomagno, T. Ligand-target interactions: what can we learn from NMR? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 245–266 (2005).

93. Lian, L. Y., Barsukov, I. L., Sutcliffe, M. J., Sze, K. H. & Roberts, G. C. K. Protein-ligand interactions: Exchange processes and determination of ligand conformation and protein-ligand contacts. *Methods Enzymol.* **239**, 657–700 (1994).
94. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
95. *Protein NMR techniques. Protein NMR Techniques* (2004).
96. Rieping, W. *et al.* ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382 (2007).
97. Guntert, P. Automated NMR protein structure calculation. **43**, 105–125 (2003).
98. Brunger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst* **54**, 905–921 (1998).
99. Turnbull, W. B. & Daranas, A. H. On the Value of *c*: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866 (2003).
100. Heller, G. T., Sormanni, P. & Vendruscolo, M. Targeting disordered proteins with small molecules using entropy. *Trends Biochem. Sci.* **40**, 491–496 (2015).
101. MicroCal™ iTC 200 System - manual.
102. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
103. Hein, M., Zilian, D. & Sotriffer, C. a. Docking compared to 3D-pharmacophores: the scoring function challenge. *Drug Discov. Today Technol.* **7**, e229–e236 (2010).
104. Kalyaanamoorthy, S. & Chen, Y.-P. P. Structure-based drug design to augment hit discovery. *Drug Discov. Today* **16**, 831–839 (2011).
105. Basse, M. J. *et al.* 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.* **41**, D824-7 (2013).
106. Hamon, V. *et al.* 2P2Ichem: focused chemical libraries dedicated to orthosteric modulation of protein-protein interactions. *Medchemcomm* **4**, 797–809 (2013).
107. Higuero, A. P., Jubb, H. & Blundell, T. L. TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database (Oxford)*. **2013**, bat039 (2013).
108. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).
109. Irwin, J. J. & Shoichet, B. K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model* **45**, 177–182 (2005).
110. Huwe, C. M. Synthetic library design. *Drug Discov. Today* **11**, 763–767 (2006).
111. Jones, G., Willett, P. & Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **245**, 43–53 (1995).
112. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
113. Nissink, J. W. M. *et al.* A new test set for validating predictions of protein-ligand interaction. *Proteins Struct. Funct. Genet.* **49**, 457–471 (2002).
114. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved



- protein–ligand docking using GOLD. *Proteins Struct. Funct. Bioinforma.* **52**, 609–623 (2003).
115. J. C. Cole, J. W. M. Nissink, R. T. *Protein-Ligand Docking and Virtual Screening with GOLD*. (Taylor & Francis CRC Press, 2005).
116. Verdonk, M. L. *et al.* Modeling Water Molecules in Protein–Ligand Docking Using GOLD. *J. Med. Chem.* **48**, 6504–6515 (2005).
117. Hartshorn, M. J. *et al.* Diverse, High-Quality Test Set for the Validation of Protein–Ligand Docking Performance. *J. Med. Chem.* **50**, 726–741 (2007).
118. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
119. Liebeschuetz, J. W., Cole, J. C. & Korb, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided. Mol. Des.* **26**, 737–748 (2012).
120. Nicolotti, O. *et al.* Strategies of multi-objective optimization in drug discovery and development. *Expert Opin. Drug Discov.* **6**, 871–884 (2011).
121. Nicolaou, C. A., Brown, N. & Pattichis, C. S. Molecular optimization using computational multi-objective methods. *Curr. Opin. Drug Discov. Devel.* **10**, 316–324 (2007).
122. Langer, T. & Wolber, G. Pharmacophore definition and 3D searches. *Drug Discov. Today Technol.* **1**, 203–207 (2004).
123. Seidel, T., Ibis, G., Bendix, F., Wolber, G. & Seidel, T. Strategies for 3D pharmacophore-based virtual screening. *Drug Discov. Today Technol.* **7**, e203-70 (2010).
124. Cole, J. C., Murray, C. W., Nissink, J. W. M., Taylor, R. D. & Taylor, R. Comparing protein–ligand docking programs is difficult. *Proteins Struct. Funct. Bioinforma.* **60**, 325–332 (2005).
125. Hevener, K. E. *et al.* Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase. *J. Chem. Inf. Model.* **49**, 444–460 (2009).
126. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist’s Guide to Molecular Interactions. *J. Med. Chem.* **53**, 5061–5084 (2010).
127. Koes, D. R. & Camacho, C. J. Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure. *Bioinformatics* **28**, 784–91 (2012).
128. Cheng, T., Li, Q., Zhou, Z., Wang, Y. & Bryant, S. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *AAPS J.* **14**, 133–141 (2012).
129. Wang, R. & Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **41**, 1422–1426 (2001).
130. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **42**, 5100–5109 (1999).
131. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
132. Berthold, M. *et al.* in *Data Analysis, Machine Learning and Applications* (eds. Preisach, C., Burkhardt, H., Schmidt-Thieme, L. & Decker, R.) 319–326 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-78246-9\_38

133. Valerie J. Gillet , Wael Khatib , Peter Willett, P. J. F. and D. V. S. G. Combinatorial Library Design Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **42**, 375–385 (2002).
134. Gleeson, M. P. Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J. Med. Chem.* **51**, 817–34 (2008).
135. Scott A. Wildman and Gordon M. Crippen. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci* **39**, 868–873 (1999).
136. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293 (1992).
137. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201–207 (1992).
138. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496–514 (1990).
139. Muhandiram, D. R. & Kay, L. E. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **103**, 203–216 (1994).
140. Grzesiek, S. & Bax, A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432–440 (1992).
141. Clubb, R. T., Thanabal, V. & Wagner, G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue <sup>1</sup>HN, <sup>15</sup>N, and <sup>13</sup>C' chemical shifts in <sup>15</sup>N <sup>13</sup>C-labelled proteins. *J. Magn. Reson.* **97**, 213–217 (1992).
142. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).
143. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11**, 852–861 (2002).
144. Richarz, R. & Wüthrich, K. Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly- X-L- Ala-OH. *Biopolymers* **17**, 2133–2141 (1978).
145. Bundi, A. & Wüthrich, K. <sup>1</sup>H-nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* **18**, 285–297 (1979).
146. Wüthrich, K. *NMR in biological research : peptides and proteins.* (North-Holland Pub. Co., 1976).
147. De Groot, C. O. *et al.* Molecular Insights into Mammalian End-binding Protein Heterodimerization. *J. Biol. Chem.* **285**, 5802–5814 (2010).
148. Habeck, M., Rieping, W., Linge, J. P. & Nilges, M. in *Protein NMR Techniques* (ed. Downing, A. K.) **278**, 379–402 (Humana Press, 2004).
149. Parry, D. a D., Fraser, R. D. B. & Squire, J. M. Fifty years of coiled-coils and  $\alpha$ -helical bundles: A close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258–269 (2008).

150. Werner., K. W. B. M. & B. Polypeptide secondary structure determination by NMR observation of short proton-proton distances. *J. Mol. Biol.* 715–740 (1984).
151. Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10714–9 (2015).
152. Nilges, M. A Calculation Strategy for the Structure Determination of Symmetric Dimers by. **309**, (1993).
153. Fossi, M., Oschkinat, H., Nilges, M. & Ball, L. J. Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data. *J. Magn. Reson.* **175**, 92–102 (2005).
154. Falchi, F., Caporuscio, F. & Recanatini, M. Structure-based design of small-molecule protein-protein interaction modulators: the story so far. *Future Med. Chem.* **6**, 343–57 (2014).
155. Kroemer, R. T. Structure-based drug design: docking and scoring. *Curr. Protein Pept. Sci.* **8**, 312–28 (2007).
156. Livingstone, D. *A Practical Guide to Scientific Data Analysis*. (John Wiley & Sons, 2009). doi:10.1002/9780470017913
157. Dundas, J. *et al.* CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, 116–118 (2006).
158. Montalbetti, C. A. G. N. & Falque, V. Amide bond formation and peptide coupling. *Tetrahedron* **61**, 10827–10852 (2005).
159. Verdonk, M. L. *et al.* Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **44**, 793–806 (2004).
160. CCDC Software. GOLD User Guide. 4, 74 (2011).
161. Ekins, S. *et al.* Analysis and hit filtering of a very large library of compounds screened against Mycobacterium tuberculosis. *Mol. Biosyst.* **6**, 2316–2324 (2010).
162. Blake, J. F. Identification and Evaluation of Molecular Properties Related to Preclinical Optimization and Clinical Fate. *Med. Chem. (Los. Angeles)*. **1**, 649–655 (2005).
163. Hann, M. M. *et al.* Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.* **39**, 897–902 (1999).
164. Huth, J. R. *et al.* ALARM NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc.* **127**, 217–224 (2005).
165. Oprea, T. I., Davis, A. M., Teague, S. J. & Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Model.* **41**, 1308–1315 (2001).
166. Baell, J. B. & Holloway, G. a. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
167. Pipeline Pilot release 8.5. (2011).
168. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug Discov. Today* **20**, 122–128 (2014).

169. Martin, S. F. & Clements, J. H. Correlating Structure and Energetics in Protein-Ligand Interactions: Paradigms and Paradoxes. *Annu. Rev. Biochem.* **82**, 267–293 (2013).
170. Pierce, M. M., Raman, C. S. & Nall, B. T. Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods* **19**, 213–221 (1999).
171. Buey, R. M. *et al.* Sequence determinants of a microtubule tip localization signal (MtLS). *J. Biol. Chem.* **287**, 28227–42 (2012).
172. Leavitt, S. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11**, 560–566 (2001).
173. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
174. Habeck, M., Rieping, W., Linge, J. P. & Nilges, M. in *Protein NMR Techniques* **278**, 3179–402 (2004).
175. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nat. Rev. Drug Discov.* (2016). doi:10.1038/nrd.2016.29
176. Hann, M. M. & Oprea, T. I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* **8**, 255–263 (2004).
177. Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **6**, 881–90 (2007).
178. Pellecchia, M., Sem, D. S. & Wüthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.* **1**, 211–219 (2002).
179. Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661–665 (1992).
180. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-Tailored Water Suppression for <sup>1</sup>H-<sup>15</sup>N HSQC Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A* **102**, 241–245 (1993).
181. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).
182. Campagne, S., Gervais, V. & Milon, a. Nuclear magnetic resonance analysis of protein – DNA interactions Nuclear magnetic resonance analysis of protein – DNA interactions. 1065–1078 (2011). doi:doi: 10.1098/rsif.2010.0543
183. Morton, C. J. *et al.* Solution structure and peptide binding of the SH3 domain from human Fyn. *Structure* **4**, 705–714 (1996).
184. Fielding, L. NMR methods for the determination of protein–ligand dissociation constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **51**, 219–242 (2007).
185. Wittekind, M. & Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B* **101**, 201–205 (1993).
186. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly <sup>13</sup>C/<sup>15</sup>N-enriched proteins. *J. Biomol. NMR* **3**, 185–204 (1993).

187. Schleucher, J., Sattler, M. & Griesinger, C. Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCO experiment. *Angew. Chem. Int. Ed. Engl.* **32**, 1489–1491 (1993).
188. Kay, L. E., Xu, G. Y. & Yamazaki, T. Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H<sub>2</sub>O Saturation. *Journal of Magnetic Resonance, Series A* **109**, 129–133 (1994).
189. Farmer, B. T., Venters, R. a., Spicer, L. D., Wittekind, M. G. & Müller, L. A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR* **2**, 195–202 (1992).
190. Hwang, T. L. & Shaka, a. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **112**, 275–279 (1995).
191. Vuister, G. W. & Bax, A. Resolution enhancement and spectral editing of uniformly <sup>13</sup>C-enriched proteins by homonuclear broadband <sup>13</sup>C decoupling. *J. Magn. Reson.* **98**, 428–435 (1992).
192. Meissner, a & Sørensen, O. W. The role of coherence transfer efficiency in design of TROSY-type multidimensional NMR experiments. *J. Magn. Reson.* **139**, 439–442 (1999).
193. Bax, A., Clore, G. M. & Gronenborn, A. M. <sup>1</sup>H-<sup>1</sup>H correlation via isotropic mixing of <sup>13</sup>C magnetization, a new three-dimensional approach for assigning <sup>1</sup>H and <sup>13</sup>C spectra of <sup>13</sup>C-enriched proteins. *J. Magn. Reson.* **88**, 425–431 (1990).
194. Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A 4D HCCH-TOCSY experiment for assigning the side chain <sup>1</sup>H and <sup>13</sup>C resonance of proteins. *J. Biomol. NMR* **2**, 655–659 (1992).
195. Kay, L. E., Xu, G. Y., Singer, a. U., Muhandiram, D. R. & Formankay, J. D. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain <sup>1</sup>H and <sup>13</sup>C Correlations in H<sub>2</sub>O Samples of Proteins. *Journal of Magnetic Resonance, Series B* **101**, 333–337 (1993).
196. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. Three-Dimensional Heteronuclear NMR of <sup>15</sup>N-labeled proteins. *J. Am. Chem. Soc.* **1111**, 1515–1517 (1989).
197. Marion, D. *et al.* Overcoming the overlap problem in the assignment of <sup>1</sup>H NMR spectra of larger proteins by use of three-dimensional heteronuclear <sup>1</sup>H-<sup>15</sup>N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to . *Biochemistry* **28**, 6150–6156 (1989).
198. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660 (1995).
199. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).
200. Cheung, M. S., Maguire, M. L., Stevens, T. J. & Broadhurst, R. W. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.* **202**, 223–233 (2010).
201. Hayashi, I., Wilde, A., Mal, T. K. & Ikura, M. Structural basis for the activation of

- microtubule assembly by the EB1 and p150Glued complex. *Mol. Cell* **19**, 449–460 (2005).
202. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315–316 (2003).
203. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes* **5**, 367 (2012).
204. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–30 (2002).
205. Fry, D. C. *et al.* Deconstruction of a nutlin: dissecting the binding determinants of a potent protein-protein interaction inhibitor. *ACS Med. Chem. Lett.* **4**, 660–665 (2013).
206. Alvarez, J. & Shoichet, B. *Drug Discovery Virtual Screening in Drug Discovery. Group 1*, (2005).
207. Kubinyi, H. QSAR and 3D QSAR in drug design. Part 1: Methodology. *Drug Discov. Today* **2**, 457–467 (1997).
208. Ooms, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.* **7**, 141–58 (2000).
209. Willett, P., Winterman, V. & Bawden, D. Implementation of nearest-neighbor searching in an online chemical structure search system. *J. Chem. Inf. Model.* **26**, 36–41 (1986).
210. Dragon (Software for Molecular Descriptor Calculation). (2014).
211. Mitchell, J. B. O. Machine learning methods in chemoinformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, n/a-n/a (2014).
212. Roy, K., Kar, S. & Das, R. N. in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* 191–229 (Elsevier Inc., 2015).  
doi:10.1016/B978-0-12-801505-6.00006-5
213. Team, R. D. C. R: a language and environment for statistical computing. (2011).
214. Pipeline Pilot - Statistics and Data Modeling: User guide. (2011).
215. Roy, K., Kar, S. & Das, R. N. in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* 231–289 (Elsevier Inc., 2015).  
doi:10.1016/B978-0-12-801505-6.00007-7
216. Sander, T., Freyss, J., Von Korff, M. & Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **55**, 460–473 (2015).
217. [ww.fortebio.com/octet-red96-specs.html](http://ww.fortebio.com/octet-red96-specs.html). Available at: [ww.fortebio.com/octet-red96-specs.html](http://ww.fortebio.com/octet-red96-specs.html). (Accessed: 4th July 2016)
218. Macarron, R. *et al.* Impact of high-throughput screening. (2011).







## APPENDIX

### A.1 Structure determination of free EB1cΔ8

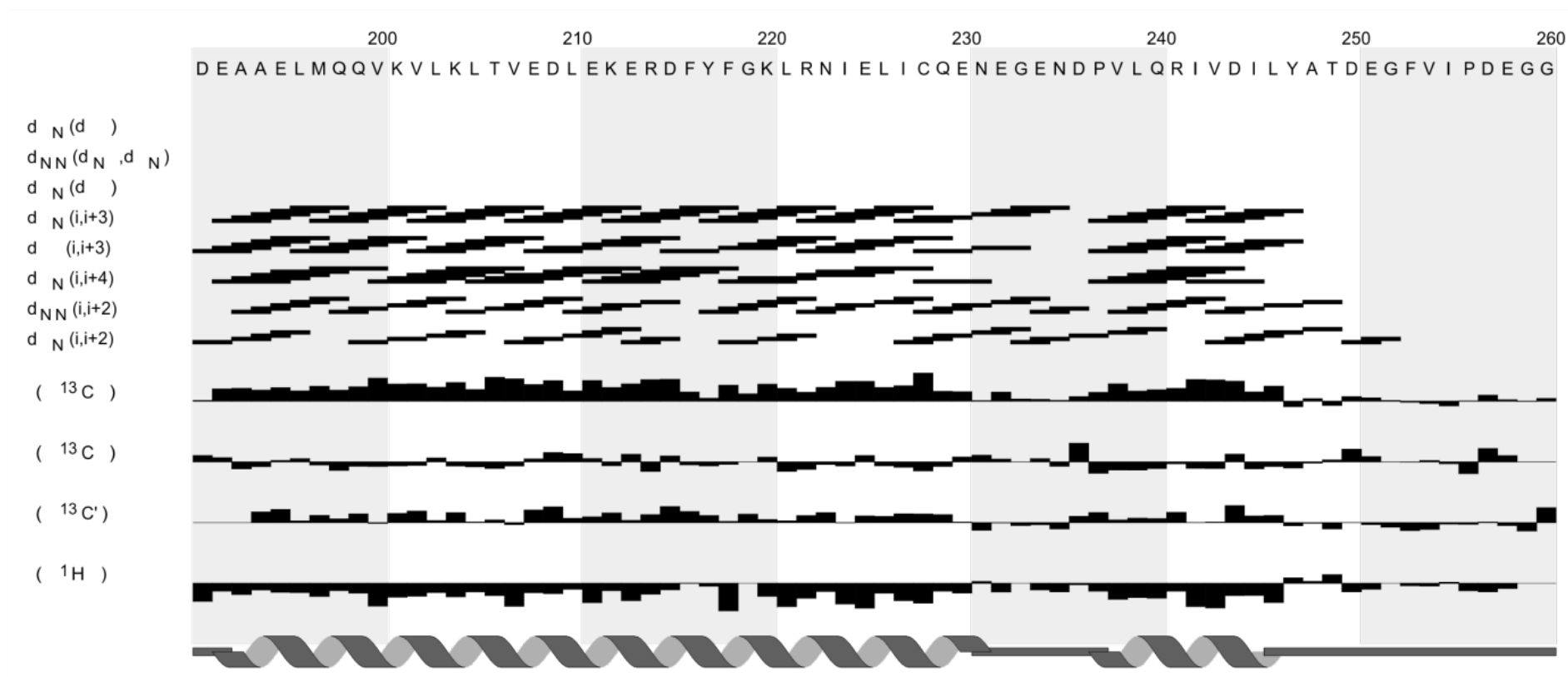


Figure 1 – Secondary structure prediction obtained from the chemical shifts of the EB1cΔ8 obtained using CcpNmr Analysis.

## A.2 Structure determination of the complex EB1cΔ8-1a

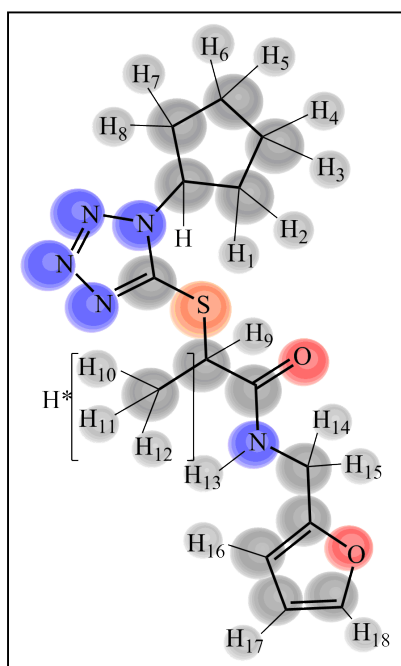


Figure 2 – structure of molecule **1a** and respective atom labels used for NMR resonance assignment.

Table 1 – intermolecular restraints between EB1cΔ8 and molecule **1a** obtained through manual assignment and directly input into ARIA.

Chain	Residue	Atom	Chain	Atom	Distance (Å)
B	Phe216	Hε	C	H*	4.8
A	Arg222	Hδ	C	H3/H5	4.1
A	Arg222	Hγ	C	H5	4.4
A	Arg222	Hγ	C	H3/H5	4.5
A	Arg222	Hδ	C	H4/H6	4.4
A	Arg222	Hδ	C	H*	4.4
A	Arg222	Hα	C	H3/H5	4.1
A	Arg222	Hγ	C	H2/H8	4.4
A	Arg222	Hδ	C	H3/H5	4.1
A	Arg222	Hα	C	H4/H6	4.4
A	Arg222	Hδ	C	H3/H5	4.1
A	Arg222	Hδ	C	H4/H6	4.4
A	Glu225	Hγ	C	H3/H5	4.6
A	Leu246	Hδ2	C	H4/H6	3.9
A	Leu246	Hδ2	C	H3/H5	4
A	Leu246	Hδ1	C	H2/H8	4.2
A	Leu246	Hδ1	C	H1	4.2
A	Leu246	Hδ1	C	H4/H6	3.4
A	Leu246	Hδ1	C	H3/H5	3.7
A	Thr249	Hγ	C	H*	3.1
A	Thr249	Hγ	C	H9	4.2

Table 2 – intermolecular restraints between EB1cΔ8 and molecule 1a obtained through the iterative assignment performed by ARIA

Chain	Residue	Atom	Chain	Atom	Distance (Å)
A	Arg222	Hδ	C	H4/H6	4.9
A	Leu246	Hδ1	C	H3/H5	3.7
A	Leu246	Hδ1	C	H4/H6	3.4
A	Leu246	Hδ2	C	H4/H6	3.8
A	Leu246	Hδ2	C	H3/H5	3.9
A	Leu246	Hδ1	C	H2/H8	4.3
A	Leu246	Hδ2	C	H1/H7	4.9
A	Leu246	Hδ2	C	H2/H8	4.6
A	Leu246	Hδ1	C	H1	4.2
A	Tyr247	Hε	C	H4/H6	4.1
A	Tyr247	Hε	C	H2/H8	3.8
A	Tyr247	Hδ	C	H3/H5	4.6
A	Tyr247	Hδ	C	H4/H6	4.3
A	Tyr247	Hδ	C	H1/H7	4.7
A	Tyr247	Hδ	C	H2/H8	4.3
A	Tyr247	Hε	C	H3/H5	3.9
A	Tyr247	Hε	C	H1/H7	4.6
A	Tyr247	Hδ	C	H*	4.2
A	Tyr247	Hδ	C	H14/H15	4.9
A	Tyr247	Hε	C	H	4.5
A	Tyr247	Hε	C	H*	4.6
A	Tyr247	Hε	C	H9	4.9
A	Thr249	Hγ2	C	H*	3.4
A	Thr249	Hγ2	C	H9	4.4

### A.3 Structure determination of the complex EB1c $\Delta$ 8-11MACF

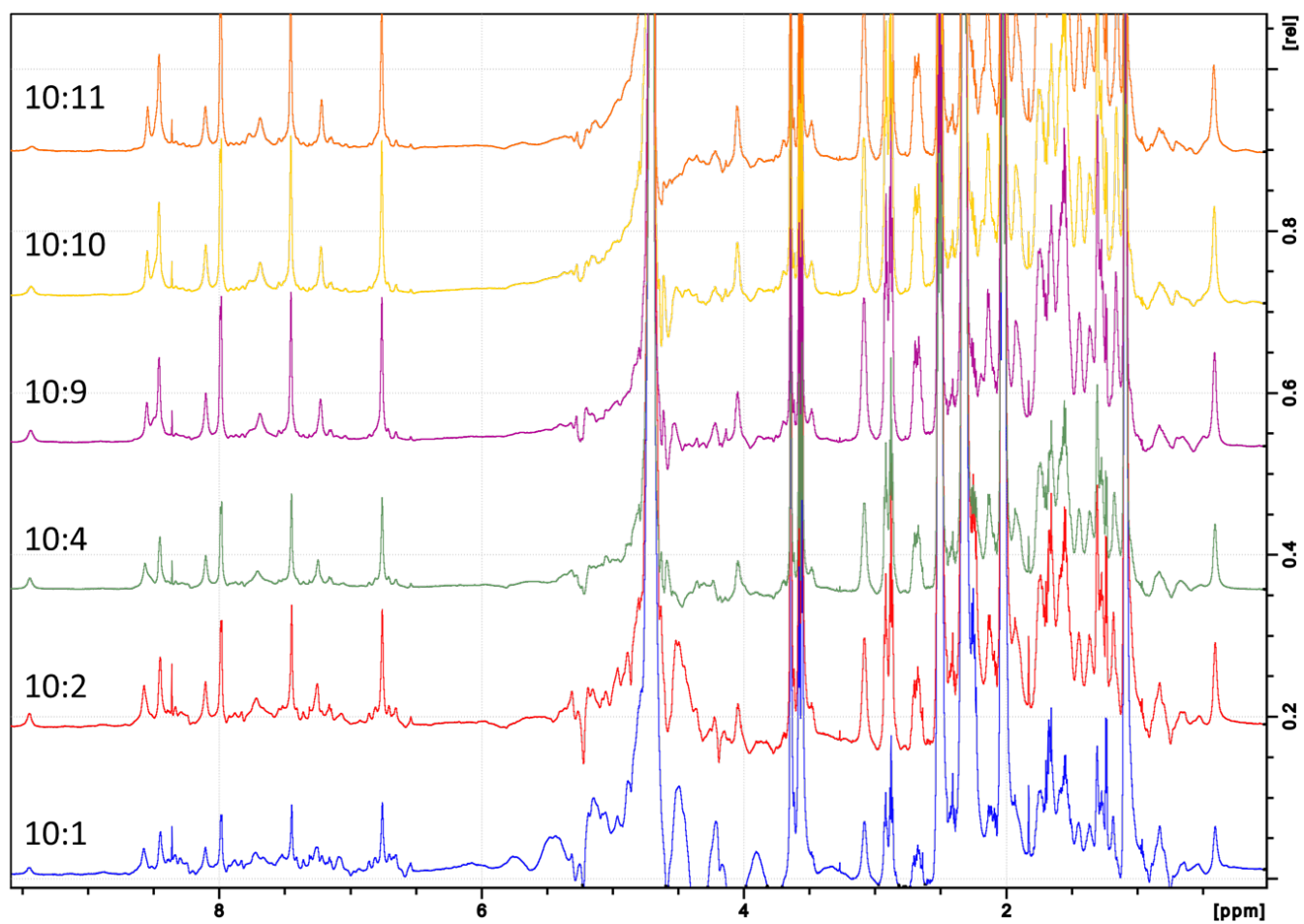


Figure 3 – 1D  $^1\text{H}$  acquired for the following protein:peptide ratios – 10:1, 10:2, 10:4, 10:9, 10:10, 10:11.

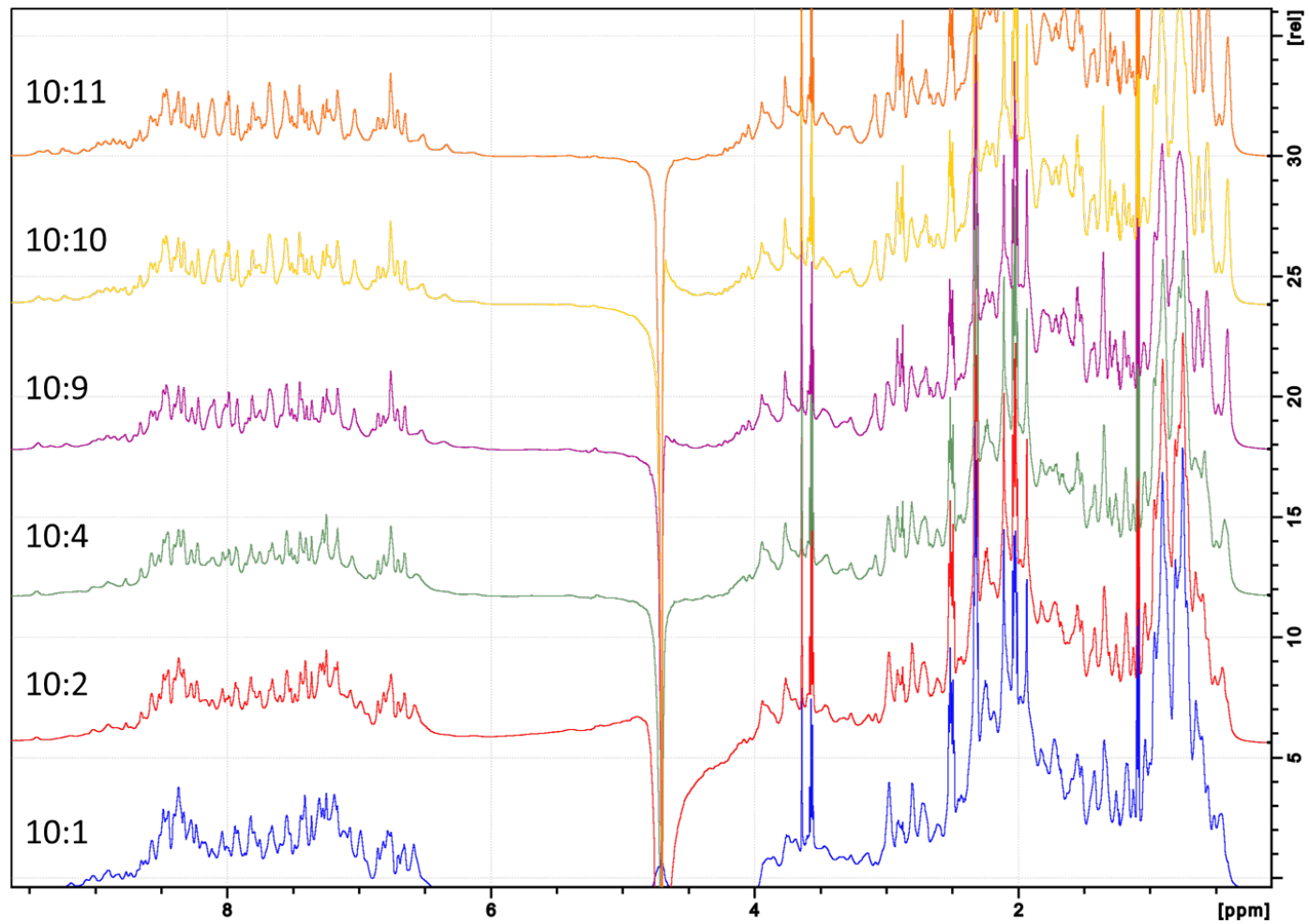


Figure 4 –  $^{15}\text{N}$ ,  $^{13}\text{C}$ -filtered-1D  $^1\text{H}$  acquired for the following protein:peptide ratios – 10:1, 10:2, 10:4, 10:9, 10:10, 10:11.

Table 3 – intermolecular restraints between EB1cΔ8 and 11MACF obtained through manual assignment and directly input into ARIA.

Chain	Residue	Atom	Chain	Residue	Atom	Distance (Å)
B	Phe216	Hζ	C	Thr5481	Hγ2	5.5
B	Phe216	Hε	C	Thr5481	Hγ2	6.4
B	Tyr217	Hε	C	Ile5479	Hγ2	4.9
B	Tyr217	Hδ	C	Ile5479	Hγ2	5.1
B	Tyr217	Hε	C	Ile5479	Hδ1	5.8
B	Tyr217	Hε	C	Ile5479	Hβ	6.1
B	Tyr217	Hδ	C	Ile5479	Hδ1	6.2
B	Tyr217	Hδ	C	Ile5479	Hβ	6.2
B	Tyr217	Hε	C	Pro5480	Hβ1	5.5
B	Tyr217	Hδ	C	Pro5480	Hβ1	5.7
B	Tyr217	Hε	C	Pro5480	Hγ2	6.2
B	Tyr217	Hδ	C	Pro5480	Hγ2	6.2
B	Tyr217	Hε	C	Pro5480	Hβ2	6.5
B	Tyr217	Hε	C	Thr5481	Hγ2	5.5
B	Tyr217	Hδ	C	Thr5481	Hγ2	5.7
A	Phe218	Hε	C	Ser5477	Hβ2	6.5
A	Phe218	Hε	C	Ile5479	Hγ2	3.8
A	Phe218	Hε	C	Ile5479	Hγ2	4.1
A	Phe218	Hδ	C	Ile5479	Hγ2	4.3
A	Phe218	Hε	C	Ile5479	Hδ1	4.4
A	Phe218	Hε	C	Ile5479	Hβ	4.5
A	Phe218	Hζ	C	Ile5479	Hγ2	4.7
A	Phe218	Hζ	C	Ile5479	Hδ1	4.9
A	Phe218	Hζ	C	Ile5479	Hβ	5.0
A	Phe218	Hδ	C	Ile5479	Hγ2	5.0
A	Phe218	Hδ	C	Ile5479	Hδ1	5.4
A	Phe218	Hδ	C	Ile5479	Hβ	5.7
A	Phe218	Hε	C	Ile5479	Hα	6.3
A	Phe218	Hε	C	Ile5479	Hγ1	6.4
A	Phe218	Hζ	C	Ile5479	Hγ1	6.5
A	Phe218	Hε	C	Pro5480	Hγ2	6.3
A	Phe218	Hε	C	Pro5480	Hβ1	6.5
A	Phe218	Hε	C	Thr5481	Hγ2	3.9
A	Phe218	Hζ	C	Thr5481	Hγ2	3.9
A	Phe218	Hε	C	Thr5481	Hγ2	4.0
A	Phe218	Hζ	C	Thr5481	Hγ2	4.4
A	Phe218	Hδ	C	Thr5481	Hγ2	4.5
A	Phe218	Hδ	C	Thr5481	Hγ2	5.1
A	Phe218	Hε	C	Thr5481	Hβ	6.3
A	Leu221	Hδ2	C	Ile5479	Hδ1	4.1
A	Leu221	Hδ2	C	Ile5479	Hγ2	4.2
A	Leu221	Hδ2	C	Ile5479	Hγ2	4.4
A	Leu221	Hδ1	C	Ile5479	Hγ2	4.5
A	Leu221	Hδ1	C	Ile5479	Hδ1	4.5
A	Leu221	Hδ2	C	Ile5479	Hδ1	4.6
A	Leu221	Hα	C	Ile5479	Hδ1	4.8
A	Leu221	Hα	C	Ile5479	Hδ1	5.0
A	Leu221	Hγ	C	Ile5479	Hδ1	5.0
A	Leu221	Hα	C	Ile5479	Hγ2	5.2

A	Arg222	H $\alpha$	C	Ile5479	H $\delta$ 1	4.3
A	Arg222	H $\alpha$	C	Ile5479	H $\delta$ 1	4.6
A	Arg222	H $\delta$ 1	C	Ile5479	H $\delta$ 1	4.9
A	Leu246	H $\delta$ 1	C	Ile5479	H $\delta$ 1	3.8
A	Leu246	H $\delta$ 1	C	Ile5479	H $\gamma$ 2	4.1
A	Leu246	H $\delta$ 2	C	Ile5479	H $\delta$ 1	4.1
A	Leu246	H $\delta$ 2	C	Ile5479	H $\gamma$ 2	4.3
A	Leu246	H $\delta$ 1	C	Ile5479	H $\delta$ 1	4.6
A	Leu246	H $\delta$ 1	C	Ile5479	H $\gamma$ 2	4.8
A	Leu246	H $\delta$ 1	C	Ile5479	H $\beta$	4.9
A	Leu246	H $\delta$ 2	C	Ile5479	H $\gamma$ 1	5.3
A	Leu246	H $\delta$ 2	C	Ile5479	H $\alpha$	5.6
A	Tyr247	H $\epsilon$	C	Lys5478	H $\beta$ 2	4.0
A	Tyr247	H $\epsilon$	C	Lys5478	H $\beta$ 1	4.2
A	Tyr247	H $\epsilon$	C	Lys5478	H $\beta$ 1	4.7
A	Tyr247	H $\epsilon$	C	Lys5478	H $\epsilon$ 2	5.1
A	Tyr247	H $\epsilon$	C	Lys5478	H $\alpha$	5.2
A	Tyr247	H $\epsilon$	C	Lys5478	H $\beta$ 2	5.4
A	Tyr247	H $\epsilon$	C	Lys5478	H $\delta$ 2	5.5
A	Tyr247	H $\epsilon$	C	Lys5478	H $\gamma$ 2	5.5
A	Tyr247	H $\epsilon$	C	Ile5479	H $\gamma$ 2	3.8
A	Tyr247	H $\epsilon$	C	Ile5479	H $\alpha$	3.9
A	Tyr247	H $\epsilon$	C	Ile5479	H $\delta$ 1	4.0
A	Tyr247	H $\epsilon$	C	Ile5479	H $\beta$	4.2
A	Tyr247	H $\epsilon$	C	Ile5479	H $\gamma$ 2	4.4
A	Tyr247	H $\epsilon$	C	Ile5479	H $\delta$ 1	4.7
A	Tyr247	H $\epsilon$	C	Ile5479	H $\beta$	5.1
A	Tyr247	H $\epsilon$	C	Ile5479	H $\alpha$	5.2
A	Tyr247	H $\delta$	C	Ile5479	H $\gamma$ 2	5.4
A	Tyr247	H $\epsilon$	C	Ile5479	H $\gamma$ 12	5.5
A	Tyr247	H $\delta$	C	Ile5479	H $\alpha$	6.0
A	Ala248	H $\beta$	C	Pro5480	H $\beta$ 1	4.8
A	Thr249	H $\gamma$ 2	C	Lys5478	H $\delta$ 1	4.6
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\beta$ 1	3.4
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\beta$ 2	3.7
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\gamma$ 2	3.7
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\alpha$	4.8
A	Thr249	H $\gamma$ 2	C	Thr5481	H $\beta$	4.7
A	Phe253	H $\epsilon$	C	Ile5479	H $\gamma$ 2	5.3
A	Phe253	H $\epsilon$	C	Ile5479	H $\delta$ 1	5.9
A	Phe253	H $\delta$	C	Thr5481	H $\gamma$ 2	5.8
A	Phe253	H $\delta$	C	Thr5481	H $\alpha$	5.8
A	Phe253	H $\epsilon$	C	Thr5481	H $\gamma$ 2	5.9
A	Phe253	H $\epsilon$	C	Thr5481	H $\alpha$	6.2
A	Phe253	H $\delta$	C	Thr5481	H $\beta$	6.3
A	Phe253	H $\epsilon$	C	Thr5481	H $\beta$	6.3
A	Phe253	H $\epsilon$	C	Pro5482	H $\gamma$ 2	5.3
A	Phe253	H $\delta$	C	Pro5482	H $\alpha$	5.6
A	Phe253	H $\epsilon$	C	Pro5482	H $\beta$ 1	5.7
A	Phe253	H $\epsilon$	C	Pro5482	H $\alpha$	5.8
A	Phe253	H $\delta$	C	Pro5482	H $\beta$ 1	5.9
A	Phe253	H $\delta$	C	Pro5482	H $\gamma$ 2	6.0
A	Phe253	H $\delta$	C	Pro5482	H $\gamma$ 1	6.1
A	Phe253	H $\delta$	C	Pro5482	H $\beta$ 2	6.1

A	Phe253	He	C	Pro5482	Hγ1	6.1
A	Phe253	He	C	Pro5482	Hβ2	6.1
A	Phe253	He	C	Pro5482	Hδ1	6.2
A	Phe253	Hδ	C	Pro5482	Hδ1	6.3
A	Phe253	He	C	Pro5482	Hδ2	6.4
A	Phe253	Hδ	C	Pro5482	Hδ2	6.6
A	Val254	Hγ1	C	Thr5481	Hγ2	4.0
A	Val254	Hγ2	C	Thr5481	Hγ2	4.5
A	Val254	Hγ1	C	Thr5481	Hγ2	4.6
A	Val254	Hγ2	C	Thr5481	Hγ2	5.3
A	Val254	Hγ1	C	Pro5482	Hα	4.0
A	Val254	Hγ2	C	Pro5482	Hα	4.1
A	Val254	Hγ2	C	Pro5482	Hβ1	4.7
A	Val254	Hγ1	C	Gln5483	Hα	3.1
A	Val254	Hγ2	C	Gln5483	Hγ2	3.1
A	Val254	Hγ1	C	Gln5483	Hα	3.1
A	Val254	Hγ2	C	Gln5483	Hα	3.3
A	Val254	Hγ1	C	Gln5483	Hγ2	3.3
A	Val254	Hγ2	C	Gln5483	Hβ2	3.3
A	Val254	Hγ2	C	Gln5483	Hγ2	3.5
A	Val254	Hγ2	C	Gln5483	Hβ1	3.5
A	Val254	Hγ1	C	Gln5483	Hγ2	3.5
A	Val254	Hγ2	C	Gln5483	Hα	3.6
A	Val254	Hγ1	C	Gln5483	Hβ2	3.6
A	Val254	Hγ1	C	Gln5483	Hβ1	3.7
A	Val254	Hγ2	C	Gln5483	Hβ2	3.7
A	Val254	Hγ2	C	Gln5483	He21	3.8
A	Val254	Hγ1	C	Gln5483	Hβ2	3.8
A	Val254	Hγ2	C	Gln5483	Hβ1	3.9
A	Val254	Hγ2	C	Gln5483	He22	3.9
A	Val254	Hγ1	C	Gln5483	Hβ1	4.0
A	Val254	Hγ2	C	Gln5483	HN	4.0
A	Val254	Hγ1	C	Gln5483	HN	4.1
A	Val254	Hγ1	C	Gln5483	He2	4.3
A	Val254	Hγ1	C	Gln5483	He22	4.5
A	Ile255	Hγ2	C	Lys5478	Hβ1	3.3
A	Ile255	Hγ2	C	Lys5478	Hα	3.4
A	Ile255	Hγ2	C	Lys5478	He2	3.6
A	Ile255	Hδ1	C	Lys5478	Hβ1	3.6
A	Ile255	Hγ2	C	Lys5478	Hδ1	3.8
A	Ile255	Hγ2	C	Lys5478	Hβ2	3.8
A	Ile255	Hδ1	C	Lys5478	He2	3.8
A	Ile255	Hδ1	C	Lys5478	Hα	3.8
A	Ile255	Hγ2	C	Lys5478	Hδ2	3.8
A	Ile255	Hγ2	C	Lys5478	He1	4.0
A	Ile255	Hδ1	C	Lys5478	Hβ1	4.0
A	Ile255	Hδ1	C	Lys5478	He2	4.1
A	Ile255	Hδ1	C	Lys5478	Hα	4.2
A	Ile255	Hδ1	C	Lys5478	Hδ1	4.2
A	Ile255	Hδ1	C	Lys5478	He1	4.2
A	Ile255	Hδ1	C	Lys5478	Hβ2	4.4
A	Ile255	Hδ1	C	Lys5478	Hδ2	4.4
A	Ile255	Hδ1	C	Lys5478	He1	4.4
A	Ile255	Hδ1	C	Lys5478	Hβ2	4.5



A	Ile255	Hδ1	C	Lys5478	Hδ2	4.6
A	Ile255	Hδ1	C	Lys5478	Hγ2	5.0
A	Ile255	Hδ1	C	Ile5479	Hα	4.7
A	Ile255	Hδ1	C	Pro5480	Hβ1	3.7
A	Ile255	Hδ1	C	Pro5480	Hβ2	4.3
A	Ile255	Hγ2	C	Pro5480	Hβ1	4.4
A	Ile255	Hδ1	C	Pro5480	Hγ2	4.4
A	Ile255	Hγ2	C	Pro5480	Hβ2	4.6
A	Ile255	Hγ1	C	Pro5480	Hβ1	4.8
A	Ile255	Hδ1	C	Pro5480	Hα	4.9
A	Ile255	Hγ1	C	Pro5480	Hβ1	5.0
A	Ile255	Hγ2	C	Pro5480	Hγ2	5.1
A	Ile255	Hγ2	C	Thr5481	Hγ2	4.3
A	Ile255	Hγ2	C	Thr5481	HN	4.7
A	Ile255	Hδ1	C	Thr5481	Hγ2	4.9
A	Ile255	Hδ1	C	Thr5481	HN	5.1
A	Pro256	Hγ2	C	Lys5478	Hα	3.6
A	Pro256	Hβ1	C	Lys5478	Hα	4.4
A	Pro256	Hγ2	C	Lys5478	Hβ1	4.4
A	Pro256	Hβ2	C	Lys5478	Hα	4.5
A	Pro256	Hδ2	C	Lys5478	Hα	4.6
A	Pro256	Hδ1	C	Lys5478	Hα	4.8
A	Pro256	Hγ2	C	Lys5478	Hδ2	5.0
A	Pro256	Hβ1	C	Lys5478	Hβ1	5.0
A	Pro256	Hγ2	C	Ile5479	Hγ2	4.6
A	Pro256	Hγ2	C	Thr5481	Hγ2	3.7
A	Pro256	Hβ2	C	Thr5481	Hγ2	4.5
A	Pro256	Hδ2	C	Thr5481	Hγ2	4.7

Table 4 – intermolecular restraints between EB1cΔ8 and 11MACF obtained through the iterative assignment performed by ARIA

Chain	Residue	Atom	Chain	Residue	Atom	Distance (Å)
A	Phe218	Hα	C	Ile5479	Hδ1	4.4
A	Leu221	Hδ2	C	Ile5479	Hγ2	3.2
A	Leu221	Hα	C	Ile5479	Hγ2	3.7
A	Leu221	Hδ2	C	Ile5479	Hδ1	3.8
A	Leu221	Hδ1	C	Ile5479	Hδ1	3.4
A	Leu221	Hα	C	Ile5479	Hδ1	3.8
A	Leu221	Hγ	C	Ile5479	Hδ1	4.4
A	Leu221	Hγ	C	Ile5479	Hδ1	4.4
A	Arg222	Hα	C	Ile5479	Hδ1	2.9
A	Arg222	Hδ1	C	Ile5479	Hδ1	4.2
A	Arg222	Hβ1	C	Ile5479	Hδ1	4.5
A	Asn223	Hα	C	Lys5475	Hγ2	3.4
A	Ile224	Hδ1	C	Ile5479	Hγ2	3.7
A	Ile224	Hγ2	C	Ile5479	Hγ2	3.9
A	Ile224	Hδ1	C	Ile5479	Hγ2	3.6
A	Ile224	Hδ1	C	Ile5479	Hδ1	3.0
A	Glu225	Hβ2	C	Ile5479	Hβ	3.6
A	Glu225	Hγ1	C	Pro5476	Hβ2	3.8

A	Glu225	H $\alpha$	C	Ile5479	H $\delta$ 1	4.1
A	Glu225	H $\gamma$ 1	C	Ile5479	H $\delta$ 1	4.3
A	Glu225	H $\gamma$ 2	C	Ile5479	H $\delta$ 1	4.5
A	Leu226	H $\delta$ 2	C	Lys5475	HE2	4.0
A	Leu226	H $\delta$ 1	C	Pro5476	H $\delta$ 2	4.1
A	Leu226	H $\delta$ 1	C	Lys5475	HE2	3.8
A	Leu226	H $\alpha$	C	Lys5475	H $\gamma$ 2	3.0
A	Leu226	H $\gamma$	C	Lys5475	H $\gamma$ 2	3.2
A	Leu226	H $\delta$ 1	C	Lys5475	H $\gamma$ 2	2.8
A	Leu226	H $\beta$ 1	C	Lys5475	H $\gamma$ 2	2.9
A	Leu246	H $\delta$ 1	C	Ile5479	H $\gamma$ 1	3.1
A	Leu246	H $\delta$ 2	C	Ile5479	H $\gamma$ 2	3.5
A	Leu246	H $\beta$ 1	C	Ile5479	H $\gamma$ 1	3.7
A	Leu246	H $\delta$ 2	C	Ile5479	H $\delta$ 1	3.0
A	Leu246	H $\alpha$	C	Ile5479	H $\gamma$ 2	3.6
A	Leu246	H $\delta$ 1	C	Ile5479	H $\gamma$ 2	4.1
A	Leu246	H $\delta$ 1	C	Ile5479	H $\delta$ 1	3.3
A	Thr249	H $\alpha$	C	Pro5480	H $\beta$ 2	4.0
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\beta$ 2	2.5
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\gamma$ 2	3.2
A	Thr249	H $\gamma$ 2	C	Pro5480	H $\beta$ 1	2.8
A	Thr249	H $\beta$	C	Pro5480	H $\beta$ 1	3.7
A	Thr249	H $\alpha$	C	Pro5480	H $\beta$ 1	3.2
A	Thr249	H $\alpha$	C	Pro5480	H $\gamma$ 2	3.3
A	Val254	H $\gamma$ 1	C	Pro5482	H $\delta$ 2	4.2
A	Val254	H $\gamma$ 1	C	Pro5480	H $\beta$ 2	3.4
A	Val254	H $\gamma$ 2	C	Gln5483	H $\beta$ 1	3.0
A	Val254	H $\gamma$ 2	C	Arg5484	H $\gamma$ 1	4.0
A	Val254	H $\gamma$ 2	C	Thr5481	H $\gamma$ 2	3.2
A	Val254	H $\gamma$ 2	C	Arg5484	HN	2.8
A	Val254	H $\gamma$ 1	C	Gln5483	H $\beta$ 1	3.2
A	Val254	H $\gamma$ 1	C	Arg5484	H $\gamma$ 1	3.9
A	Val254	H $\gamma$ 1	C	Thr5481	H $\gamma$ 2	3.7
A	Val254	H $\gamma$ 1	C	Thr5481	HN	2.8
A	Val254	H $\gamma$ 1	C	Gln5483	HE2	3.9
A	Val254	H $\beta$	C	Thr5481	HN	3.2
A	Val254	H $\beta$	C	Pro5482	H $\gamma$ 2	3.8
A	Val254	H $\gamma$ 2	C	Gln5483	H $\beta$ 1	4.1
A	Val254	H $\gamma$ 1	C	Gln5483	H $\alpha$	2.4
A	Val254	H $\gamma$ 1	C	Thr5481	H $\gamma$ 2	3.7
A	Val254	H $\gamma$ 1	C	Thr5481	H $\beta$	3.8
A	Val254	H $\gamma$ 2	C	Gln5483	H $\alpha$	2.9
A	Val254	H $\gamma$ 2	C	Gln5483	H $\gamma$ 2	2.8
A	Val254	H $\gamma$ 2	C	Gln5483	H $\beta$ 2	2.7
A	Val254	H $\gamma$ 2	C	Gln5483	HE2	3.2
A	Val254	H $\gamma$ 2	C	Gln5483	HE2	3.3

A	Val254	H $\gamma$ 1	C	Gln5483	H $\gamma$ 2	3.0
A	Val254	H $\gamma$ 1	C	Gln5483	H $\beta$ 2	3.0
A	Val254	H $\gamma$ 1	C	Gln5483	HN	3.3
A	Val254	H $\gamma$ 1	C	Gln5483	HE2	3.8
A	Val254	H $\gamma$ 2	C	Pro5482	H $\alpha$	3.6
A	Val254	H $\gamma$ 1	C	Pro5482	H $\alpha$	3.4
A	Val254	H $\gamma$ 2	C	Thr5481	H $\beta$	3.8
A	Ile255	H $\gamma$ 2	C	Pro5480	H $\beta$ 2	4.5
A	Ile255	H $\beta$	C	Pro5480	H $\beta$ 1	5.5
A	Ile255	H $\alpha$	C	Thr5481	HN	3.2
A	Ile255	H $\gamma$ 1	C	Pro5480	H $\beta$ 1	3.7
A	Ile255	H $\delta$ 1	C	Pro5480	H $\beta$ 1	3.0
A	Ile255	H $\delta$ 1	C	Lys5478	H $\alpha$	3.3
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\gamma$ 2	3.5
A	Ile255	H $\delta$ 1	C	Lys5478	H $\delta$ 1	3.8
A	Ile255	H $\delta$ 1	C	Ile5479	H $\gamma$ 2	3.8
A	Ile255	H $\gamma$ 2	C	Lys5478	HE2	3.1
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\beta$ 1	2.8
A	Ile255	H $\delta$ 1	C	Lys5478	HE2	3.3
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\delta$ 1	2.6
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\alpha$	2.4
A	Ile255	H $\gamma$ 2	C	Pro5480	H $\gamma$ 2	4.2
A	Ile255	H $\gamma$ 2	C	Pro5480	H $\beta$ 2	3.8
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\beta$ 2	3.2
A	Ile255	H $\gamma$ 1	C	Pro5480	H $\beta$ 1	3.5
A	Ile255	H $\delta$ 1	C	Pro5480	H $\alpha$	3.9
A	Ile255	H $\delta$ 1	C	Pro5480	H $\gamma$ 2	3.7
A	Ile255	H $\delta$ 1	C	Pro5480	H $\beta$ 2	3.5
A	Ile255	H $\gamma$ 2	C	Lys5478	H $\delta$ 2	3.8
A	Pro256	H $\delta$ 2	C	Thr5481	H $\gamma$ 2	3.9
A	Pro256	H $\delta$ 2	C	Lys5478	H $\alpha$	3.0
A	Pro256	H $\gamma$ 2	C	Lys5478	H $\delta$ 1	3.5
A	Pro256	H $\gamma$ 2	C	Ile5479	H $\gamma$ 2	3.5
A	Pro256	H $\gamma$ 2	C	Lys5478	H $\gamma$ 2	3.5
A	Pro256	H $\beta$ 1	C	Lys5478	H $\gamma$ 2	3.8
A	Pro256	H $\delta$ 2	C	Thr5481	HN	3.7
A	Pro256	H $\delta$ 1	C	Thr5481	H $\gamma$ 2	4.1
A	Pro256	H $\alpha$	C	Thr5481	H $\gamma$ 2	3.8
A	Pro256	H $\gamma$ 2	C	Lys5478	H $\beta$ 2	3.6
A	Pro256	H $\gamma$ 2	C	Lys5478	H $\beta$ 1	3.4
A	Pro256	H $\gamma$ 2	C	Thr5481	H $\gamma$ 2	3.5
A	Pro256	H $\gamma$ 2	C	Thr5481	HN	3.7
A	Pro256	H $\beta$ 2	C	Thr5481	H $\gamma$ 2	3.6
A	Pro256	H $\gamma$ 2	C	Lys5478	H $\alpha$	2.3
A	Asp257	H $\beta$ 2	C	Lys5478	H $\alpha$	2.9
A	Asp257	H $\beta$ 1	C	Lys5478	H $\alpha$	3.0

## A.4 Screening of a PPI focused library

It has been shown that successful inhibitors of protein– protein interactions tend to have certain properties that distinguish them from more traditional drugs designed for more conventional target classes, such as enzymes, G protein-coupled receptors, ion channels, nuclear hormone receptors, *etc.*<sup>20</sup>> It has been shown that these compounds tend to be larger and more three-dimensional.<sup>205</sup> A more detailed description of the properties identified for this class of molecules has been detailed in Chapter 1, section 1.2.1.3.

Over the past few years there have been efforts in building chemical libraries enriched in small molecules with these desirable properties. A compound library containing 1534 compounds reported to have chemical properties within the chemical space of PPI modulators was made available during the course of this project – PPI-Net compound library. The library was obtained through the “Protein-Protein Interactions Network” (PPI-Net) and funded by EPSRC grants (EP/I037210/1 and EP/I037172/1) and GlaxoSmithKline (GSK).

The strategy applied to screen this library and described herein is different from what was performed in the previous chapters but it is a valid complementary approach for identification of PPI modulators.

### A.4.1 Virtual screening of the PPI-Net library

The entirety of the PPI-Net library (1534 molecules) was subjected to an initial *in silico* screen using a structure-based virtual screening approach – docking. For that purpose GOLD<sup>111–118</sup> was used. An important use of protein–ligand docking programs is virtual screening of large libraries where each compound in turn is docked into a target binding site and scored for their predicted strength of binding. Ideally, the docking needs to be quick but the results reliable. Speeding up a docking protocol is often done at the cost of sampling fewer binding modes, which can lead to reduced success rates.<sup>114</sup> The default genetic algorithm (GA) parameters in GOLD result in thorough but relatively slow searches *per* processing core, acceptable for individual ligand docking but not of much use for large libraries, unless many cores are used in parallel.<sup>20</sup>> In virtual high throughput screening (HTS) it is more important to predict relative ligand binding affinities than accurate binding poses. Therefore, the choice of

scoring function is clearly important. It has been reported that Goldscore outperforms Chemscore in predicting affinities, especially at fast search settings. In addition, the *search efficiency* (from GA settings) can be used to further control the speed of docking and the predictive accuracy (*i.e.* the predictivity) of the results. With the *search efficiency* set at 100% GOLD will attempt to apply optimal settings for each ligand. For a ligand with five rotatable bonds this will be around 30000 GA operations. If the *search efficiency* is set to 50%, then GOLD will perform around 15000 operations thereby speeding up the docking by a factor of two, but the search space would be less well explored. The *virtual screening* option offered by GOLD sets the *search efficiency* at 30%. This setting is suitable for routine work and usually gives comparable predictive accuracy to the slower settings, unless the ligand has a large number of rotatable torsions.<sup>160</sup> Therefore, the 1534 compounds were screened using this option.

After structure-based screening it is necessary to perform a virtual hit compound selection. There are several ways of achieving this, and some of these were described in the previous chapters, *e.g.* ligand efficiency, Pareto ranking, *etc.* In order to complement these approaches and as there was already some information about which molecules formed favourable interactions with the target protein – we used quantitative structure-activity relationship (QSAR) analysis was complemented with compound plate selection. These approaches will be subsequently described.

#### **A.4.2 Quantitative structure-activity relationship (QSAR) analysis**

Drug discovery is an iterative process with a cycle of design, synthesis and testing. Analysis of the results from one iteration provides information and knowledge that will be introduced in the next cycle, enhancing it and yielding more results.<sup>71</sup> At this stage of the project there were a small number of molecules tested, some with “positive” data (*i.e.* interaction with EB1) and some with “negative” data. As described in the introductory chapter – section 1.2.1.2.1 – when a number of active and inactive compounds are known it is possible to perform a QSAR analysis. QSAR analysis refers to the use of machine learning methods to correlate structural or physiochemical properties (molecular descriptors) of a molecule with a measured property, such as biological or chemical reactivity.<sup>71</sup> Classical QSAR methods rely on structural features or lipophilicity, polarizability, electronic and steric properties (Hansch analysis). With

the advance of structural biology methods such as X-ray protein crystallography, additional information from the three dimensional structures could be integrated into the QSAR models – 3D QSAR. The comparative molecular field analysis – CoMFA method was the first real 3D QSAR method. A CoMFA analysis identifies regions in three-dimensional space that are favourable or unfavourable for a ligand-target interaction.<sup>20</sup> A group of chemically related compounds with the same mechanism of action are selected. In contrast with classical QSAR methods these should share a common pharmacophore and not the same molecular scaffold.<sup>207</sup>

Norfloxacin is one of the first examples of success of application of QSAR methods. This scaffold was found through 2D-QSAR studies made on the antibacterial activities of monosubstituted 1-ethyl-1,4-dihydro-4-oxo-quinoline-3-carboxylic acids - Figure . Donepezil hydrochloride is another successful application of QSAR methods, where X-ray crystallography and molecular modelling studies were used in combination with QSAR analysis - Figure .<sup>20</sup>

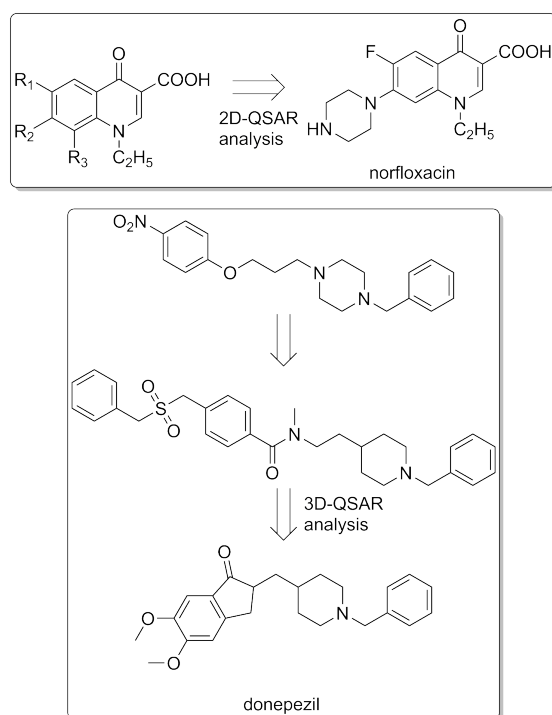
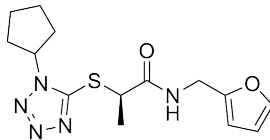
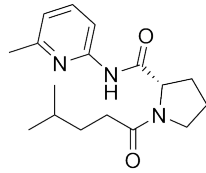
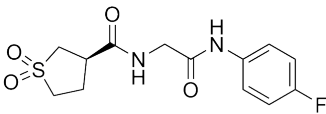
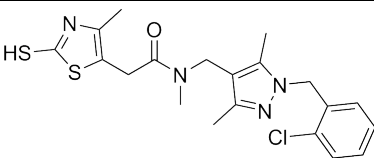
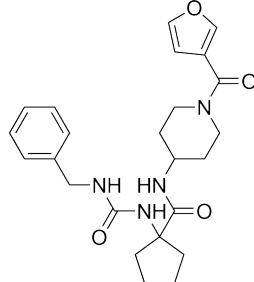
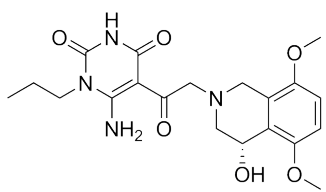


Figure 1 – Examples of molecules currently used as drug that were found through the successful application of QSAR methods.

The first step in the QSAR analysis is to compile all the tested compounds against EB1cΔ8, and classify them as active/inactive – Table 1. Compounds that promoted significant NMR chemical shift changes in the NH resonance of EB1cΔ8 are

considered as active and therefore have a green mark. Inactive compounds that did not promote chemical shift changes in NH resonances of EB1cΔ8 are marked as red crosses. The first generation compounds correspond to the virtual screening and *in vitro* testing described in Chapter 2. Compound ZINC12677264 (**1d**) is not included since at the time of this analysis the compound had not been tested. Second generation compounds correspond to the work described in Chapter 4. Finally, in-house compounds, correspond to a small compound collection existent prior to the start of this project that was screened using the same method used for the previous compounds – chemical shift perturbations upon ligand titration.

Table 1 – Molecules previously tested against EB1cΔ8.

Code	Structure	Binding to EB1cΔ8	
ZINC08389070		✓	First generation
ZINC71025726		✗	First generation
ZINC64398049		✗	First generation
ZINC40017212		✓	Second generation
ZINC31040053		✓	Second generation
ZINC12929029		✗	Second generation

ZINC50136319		X	In-house compounds
ZINC51305175		X	In-house compounds
ZINC56787233		X	In-house compounds
ZINC64898686		X	In-house compounds
ZINC66121638		X	In-house compounds
ZINC02638973		X	In-house compounds
ZINC67146994		X	In-house compounds
ZINC67208611		X	In-house compounds
ZINC67846598		X	In-house compounds
ZINC67962567		X	In-house compounds
ZINC68228234		X	In-house compounds
ZINC68451427		X	In-house compounds



For the QSAR ligand-based design it is necessary to calculate a set of molecular descriptors which capture numerically the structure and properties of each compound. These descriptors vary in complexity, from molecular weight to quantum mechanical derived values, and their complexity is usually associated with a better discriminatory capability.<sup>71</sup>

#### **A.4.3 QSAR descriptors**

The descriptors selected cover not only properties intrinsically related with the molecular structure (ligand based) but also with the predicted binding mode for each compound – docking based descriptors. These are schematically represented in Figure 2 and detailed in Chapter 7 – section 7.1.4.

The 2D descriptors are based on the two-dimensional and three-dimensional structure of the molecules, whereas 3D descriptors are based on the conformation of the molecule in its predicted binding mode. Fingerprints and physicochemical properties were both calculated using Pipeline Pilot<sup>16</sup>. Molecular fingerprints define the structure of a molecule using a series of binary digits (bits) that represent the presence or absence of particular substructures in the molecule.<sup>209</sup> Dragon descriptors were calculated using Dragon 6<sup>210</sup>, a software that can calculate up to 4885 molecular descriptors. Given the large number of molecular descriptors available, selection procedures can be used to identify those descriptors which best represent the data. Some molecular descriptors calculated using Dragon 6 correspond to molecular properties and therefore overlap with physicochemical properties (*e.g.* molecular weight). Finally, the three-dimensional descriptors, based on binding prediction and derived from the docking results can be divided in two groups: based on GOLD solutions and descriptors calculated through GoldMine based on docking predictions. In the first group one can encounter descriptors such as predicted binding score, scores from hydrogen bond formation between the ligand and the protein, *i.e.*, the output present in a Gold solution file. The second group includes the descriptors calculated in GoldMine based on the docking output. Examples are number of hydrogen bond contacts, hydrophobic buried area, number of all contacts between the ligand and the protein (general contact count), *etc.* Both sets of 3D descriptors are detailed in Table 2 and 3.

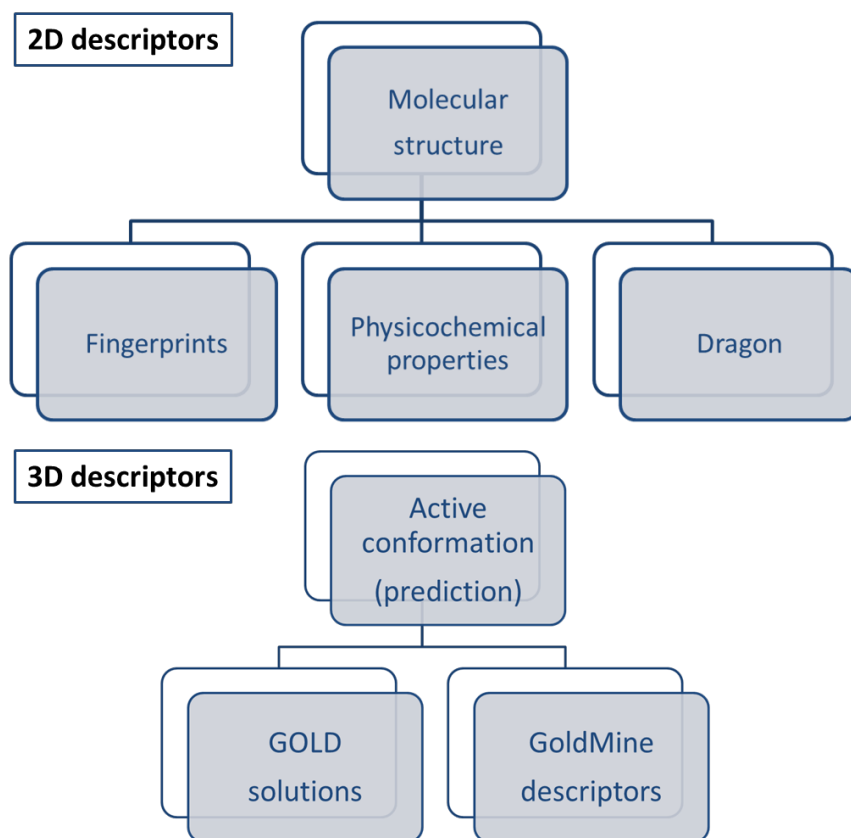


Figure 2 – Types of molecular descriptors calculated on a set of 18 molecules previously tested against EB1cA8. Top scheme shows the descriptors calculated only based on the chemical structure of the molecules. Bottom scheme represents the two sets of descriptors calculated based on the docking results – these include parameters relative to the molecular structure of the ligands but also regarding the pose obtained from docking calculations.

Table 2 – Molecular descriptors obtained directly from GOLD docking solutions.

ASP	Gold.ASP.ASP	Gold.ASP.Internal.Correction
	Gold.ASP.DEClash	Gold.ASP.Map
	Gold.ASP.DEInternal	Gold.ASP.SBar
	Gold.ASP.Fitness	
Chemscore	Gold.Chemscore.DEClash	Gold.Chemscore.Internal.Hbond
	Gold.Chemscore.DEClash.Weighted	Gold.Chemscore.Internal.Hbond.Weighted
	Gold.Chemscore.DEInternal	Gold.Chemscore.Lipo
	Gold.Chemscore.DEInternal.Weighted	Gold.Chemscore.Lipo.Weighted
	Gold.Chemscore.Fitness	Gold.Chemscore.Metal
	Gold.Chemscore.Hbond	Gold.Chemscore.Metal.Weighted
	Gold.Chemscore.Hbond.Weighted	Gold.Chemscore.SBar
	Gold.Chemscore.Internal.Correction	Gold.Chemscore.Zero.Coeff
	Gold.Chemscore.Internal.Correction.Weighted	
Goldscore	Gold.Goldscore.External.Hbond	Gold.Goldscore.Internal.Hbond
	Gold.Goldscore.External.Hbond.Weighted	Gold.Goldscore.Internal.Hbond.Weighted
	Gold.Goldscore.External.Vdw	Gold.Goldscore.Internal.Torsion
	Gold.Goldscore.External.Vdw.Weighted	Gold.Goldscore.Internal.Torsion.Weighted
	Gold.Goldscore.Fitness	Gold.Goldscore.Internal.Vdw.Weighted
	Gold.Goldscore.Internal.Correction	
	Gold.Goldscore.Internal.Correction.Weighted	
ChemPLP	Gold.PLP.Chemscore.CHOScore	Gold.PLP.ligand.clash
	Gold.PLP.Chemscore.Hbond	Gold.PLP.ligand.torsion
	Gold.PLP.Chemscore.Internal.Correction	Gold.PLP.part.buried
	Gold.PLP.Chemscore.Metal	Gold.PLP.part.hbond
	Gold.PLP.Fitness	Gold.PLP.part.metal
	Gold.PLP.PLP	Gold.PLP.part.nonpolar
	Gold.PLP.SBar	Gold.PLP.part.repulsive

Table 3 – Molecular descriptors calculated using GoldMine for solutions obtained for each scoring function, ASP, Chemscore, Goldscore and ChemPLP.

Contact descriptors	General contact count
	Hydrogen bond contact count
	Hydrophobic buried area
	Hydrophobic contacts contact count
Simple descriptors	Exposed hydrophobic count
	Ligand acceptor count
	Ligand clash count
	Ligand donor count
	Ligand hbond count
	Molecular weight
	Occluded ligand acceptor count
	Occluded ligand donor count
	Occluded ligand polar count
	Occluded protein acceptor count
	Occluded protein donor count
Occluded protein polar count	
Occluded rotatable bond count	
Surface Area Ligand Buried Surface Area	Docked
	Undocked

Given the large amount of information it is likely that some of this information may be redundant, therefore, descriptor selection is an integral part of QSAR development. It should help to reduce the vast number of descriptors, increasing the chance of finding a significant QSAR model, *i.e.* only a small number descriptors are actually correlated with biological activity.<sup>71,20</sup> The first type of descriptors to remove will be the constant values, as nothing is gained from their inclusion. Correlated descriptors should be also removed. Lastly, they should be evenly distributed and without outliers. This may require scaling some descriptors.<sup>71</sup> For example, from the descriptors calculated for Goldscore from the solution output file 12 descriptors were removed, such as “protein active residues”, common to all the docking runs and therefore contained no discriminatory information, keeping a final number of 12 used descriptors.

#### **A.4.4 Method validation**

With the first step of QSAR analysis, encoding molecular properties into numeric values, accomplished the next step was to build a predictive model. Random Forest (RF) and Support Vector Machine (SVM) were used as machine learning methods to build the predictive models. RF is based on an ensemble, or forest, of decision trees. The trees are built using a training sample of a reduced size selected at random with replacement from original data. Using the new training sample, a tree is grown with randomly selected descriptors. The remaining training data is used to estimate error and variable importance. RF is easy to use, as the user needs to fix only two parameters: the number of trees in the forest and the number of descriptors in each tree. A large number of trees should be grown, and the number of descriptors to be taken is the square root of the total number of descriptors.<sup>211,21</sup> SVM maps the data into a high-dimensional space, using a kernel function that is typically nonlinear. An optimal separation between two classes is achieved by maximising the margin between the closest points, known as support vectors.<sup>211,212</sup> Both seem to perform well in terms of classification models, performing similarly when compared. These methods have been widely used in chemoinformatics, and are easily accessible via platforms such as  $R$ <sup>213</sup>, making them a good starting point for machine learning methods.

One of the dangers in QSAR analysis is building a model that lacks both explanatory and predictive ability, being essentially meaningless. Because of this risk, it is

important to validate the built models and, if necessary, make adjustments to methods or parameters to improve their quality.<sup>21></sup> Therefore, the resulting datasets were validated using cross validated receiver operator curve area under the curve (ROC AUC) score. A cross-validation method excludes a fraction of the training data from the model building process and makes predictions for the left-out data using the model built from the remaining data. ROC curve is a visual illustration of the success and error observed in a classification model. A perfect model will score 1.0 meaning that there is a 100% true positive rate and a 0% false positive rate. When a score is 0.5 the test becomes useless as it makes a random prediction, 50% true positive versus 50% false positive, meaning that a “useful” ROC score needs to be between 0.5 and 1.<sup>215</sup> Based on the ROC scores, from the descriptors mentioned in Table 4, only the ones calculated based on the binding pose (GOLD) were used for analysis as they proved to be able to perform a better discrimination in both machine learning models (ROC >> 0.5). Therefore, three models were calculated for each scoring function – one based on the parameters calculated for each docking solution (*i.e.* solutions), the second was based on descriptors calculated using GoldMine (*i.e.* calculated descriptors) but not the direct output from the docking. A third model that combines both sets of descriptors was also calculated (solutions and calculated descriptors).

Most of the so-called classification models are really ranking models. That is, their basic prediction is not a specific class assignment but instead a score (or set of scores) that indicates the relative likelihood that a sample is in one class as opposed to another.<sup>21></sup> To turn a ranker into a true classifier requires specification of a cut-off value for the score. For score values below the cut-off, one classification is attributed (*e.g.* inactive); for scores above the cut-off they are classified alternatively (*e.g.* active). The cut-off value chosen for a given model depends on our relative preference for true positives versus false positives, or specificity versus sensitivity. In this case, molecules with a score predicted to be higher than 0.5 were classified as active.

Table 4 – ROC scores obtained for cross validation of the models built based on each set of descriptors as an evaluation of the applicability of each set of descriptors and model.

Descriptors	Random Forest ROC score	Support Vector Machine ROC score
<b><i>Molecular Fingerprints</i></b>		
FCFP_2	0.60	0.60
FCFP_4	0.64	0.51
FCFP_6	0.62	0.56
ECFP_2	0.60	0.82
ECFP_4	0.78	0.56
ECFP_6	0.58	0.51
Estate keys	0.93	0.51
MDL public keys	0.56	0.62
<b><i>Physicochemical descriptors</i></b>		
Pipeline Pilot	0.76	0.60
<b><i>GOLD</i></b>		
Chemscore – solutions	0.99	0.99
Chemscore – GoldMine descriptors	0.99	0.99
Chemscore - solutions & GoldMine descriptors	0.99	0.99
Goldscore – solutions	0.99	0.99
Goldscore – GoldMine descriptors	0.99	0.99
Goldscore - solutions & GoldMine descriptors	0.99	0.99
ASP – solutions	0.99	0.97
ASP – GoldMine descriptors	0.99	0.99
ASP – solutions & GoldMine descriptors	0.99	0.99
ChemPLP – solutions	0.99	0.97
ChemPLP – GoldMine descriptors	0.99	0.99
ChemPLP - solutions & GoldMine descriptors	0.99	0.99
<b><i>Dragon descriptors</i></b>		
2D	0.67	0.56
3D	0.64	0.56

Finally, and because not all molecules may be suitable to be evaluated using these models the *model applicability filter* node of Pipeline Pilot was used to remove molecules that are outside the models' applicability domain.<sup>21</sup> Because the model was built on a training set, molecules outside the chemical space defined by the training set are not suitable to be screened using these models. A total of 208 (out of 1534) molecules passed the *model applicability filter*, from which only 49 had

solubility values (logS) higher than -4. As reported in previous chapters (Chapters 2 and 4) -4 is an acceptable cut off value for aqueous solubility so the molecules can be screened using biophysical methods such as solution NMR and ITC. These were ranked using a *Pareto ranking* approach based on the probability of a molecule being active calculated by each model and aqueous solubility. Next section will describe this analysis.

#### A.4.5 Pareto rank

As described above three models based on descriptors obtained from docking solutions, GOLD solutions, GoldMine descriptors and both, were built for each scoring function and for each machine learning method – SVM and RF. Each model defined the molecules in terms of the probability of being active (true) or inactive (false) and attributed a score to each molecule. The molecules were then ranked using a multiobjective analysis – Pareto ranking.

A Pareto ranking node was created for each model based on descriptors obtained from the docking using each scoring function – Goldscore, Chemscore, ChemPLP and ASP. Molecules with a higher probability to be active were ranked higher, followed by aqueous solubility ranking (higher the value of log S the better) – Figure 3.

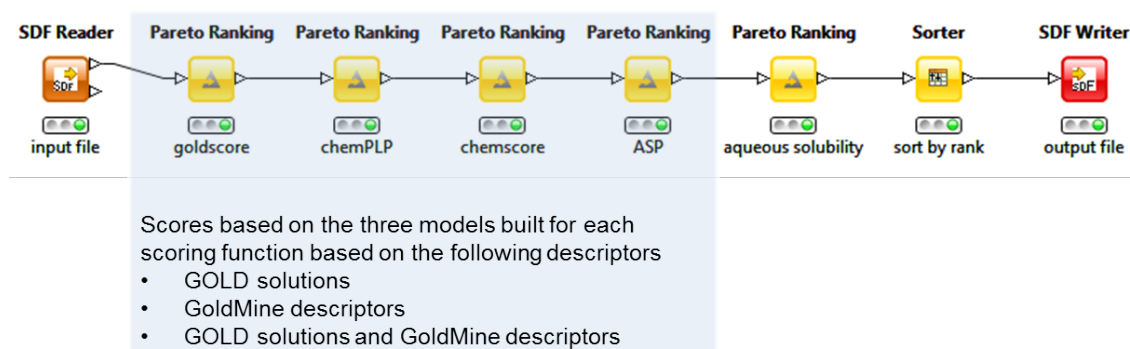


Figure 3 – Scheme describing the Pareto ranking approach used to rank the compounds from the PPI-Net library using the scores obtained from the QSAR models.

This approach was adopted for both SVM based and RF based models. As mentioned before a total of 49 molecules passed the model applicability filter and had a solubility values higher than -4, and therefore only these molecules were ranked in this manner. Analysis of the ranks obtained for both RF and SVM models reveals that for the ten best ranked molecules, eight are common to both ranks – Table 5. The presence of aromatic rings is high, 17 in total from which four are heteroaromatic.

Table 5 – Ten best ranked compounds for both RF and SVM models.

ID	RF rank	SVM rank	Structure	ID	RF rank	SVM rank	Structure
777	1	1		1219	7	8	
1428	2	2		871	-	8	
367	3	3		454	8	-	
675	4	6		603	9	-	
1372	5	5		661	-	9	
75	6	10		668	10	4	



Three of the compounds have macrocycles, between seven and eight membered rings. In terms of solubility values these vary between -3.08 and -3.96, still within an ideal range for biophysical methods (NMR and ITC) testing.

At this time there was the opportunity of testing a large number of compounds using a HTS method - Octet® RED96 System. Consequently, an alternative virtual screening method was used in order to maximise the number of compounds with the highest probability of success tested in a short period of time. This method will be subsequently described.

#### **A.4.6 Selection of compounds for high throughput screening method - Octet® RED96 System**

Virtual screening has been largely mentioned in this thesis as a way of ranking a library of compounds from best to worst and it also can be used to facilitate the biological screening of the compounds, and one of the ways described herein has been related with the choice of molecules with adequate aqueous solubility values.

When screening a large library of compounds one must be aware of the time consuming step of sample preparation for biological screen. In principle, sample preparation of a whole 96 well plate is quicker than selecting samples from 96 wells from different plates. Therefore, it was decided to develop a virtual screening method that will indicate the 96-well plate(s) out of 20 which contains the highest number of virtual hits.

To start, all compounds were docked, using GOLD, and two methods used were normal docking and biased docking. These are more detailed in the methodology chapter - Chapter 7. At this stage some important binding features of EB1 were already known and the bias of docking studies can improve the hit rate by introducing some restraints that will rank higher compounds establishing desired contacts, occupying desired positions.<sup>15</sup> Specifically, the two restraints included at this stage were a region constraint and a hydrogen bond constraint. The first will rank higher molecules that pose in a way that hydrophobic atoms are within a region defined as 3.5 Å region around Leu221 and Leu246, since these two residues are part of the hydrophobic cavity thoroughly characterised in Chapter 3. The second will favour scaffolds that make a hydrogen bond contact with the backbone of Val254 (Figure 4), also mentioned in this

thesis as a fundamental contact with the C-terminus of EB1c. This aspect is mostly addressed in Chapter 5. The reasoning for the use of these two constraints is to give higher scores to compounds that, when bound to EB1, make hydrophobic contacts with Leu221/Leu246 and/or hydrogen bond contacts with Val254, and are therefore, similar, to the native interactions.

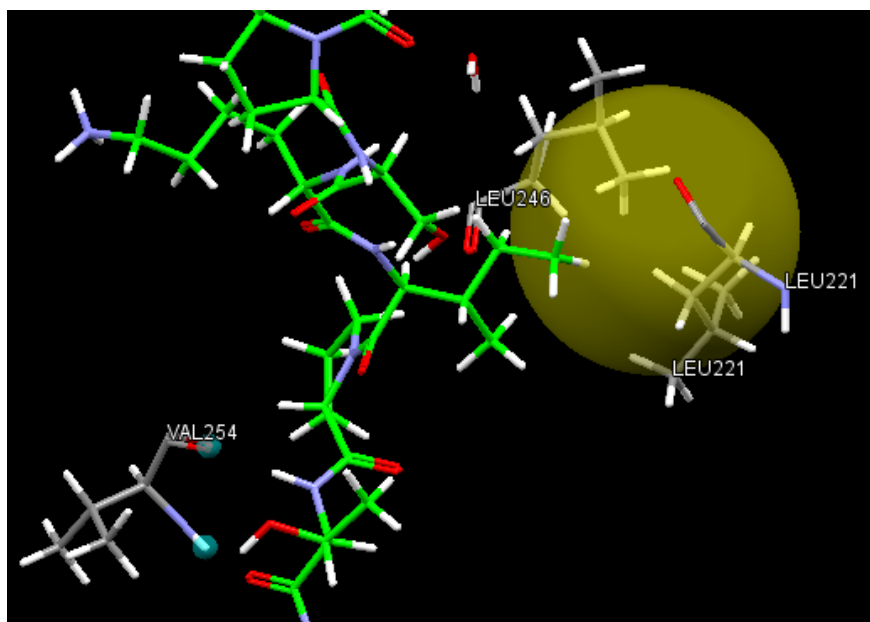


Figure 4 – Constraints used for the biased docking method. The yellow sphere represents a 3.5 Å radius around the coordinates -19.61, -26.09, 37.19. The sphere was designed in order to accommodate most of leucine side chains for Leu221 and Leu2446. The two green spheres represent hydrogen bond constraints with atoms number 515 and 518, corresponding to NH and CO of Val254 respectively.

Therefore, two different docking studies were performed, both using Goldscore as scoring function. Because the first docking study was performed without any docking constraint will be termed as “normal”, whereas the second with the two described constraints will be termed as “biased”. Subsequently, the poses obtained for each docking study using the GA and scored using Goldscore, were rescored using the three alternative scoring functions: ASP, Chemscore and ChemPLP – Figure 5.

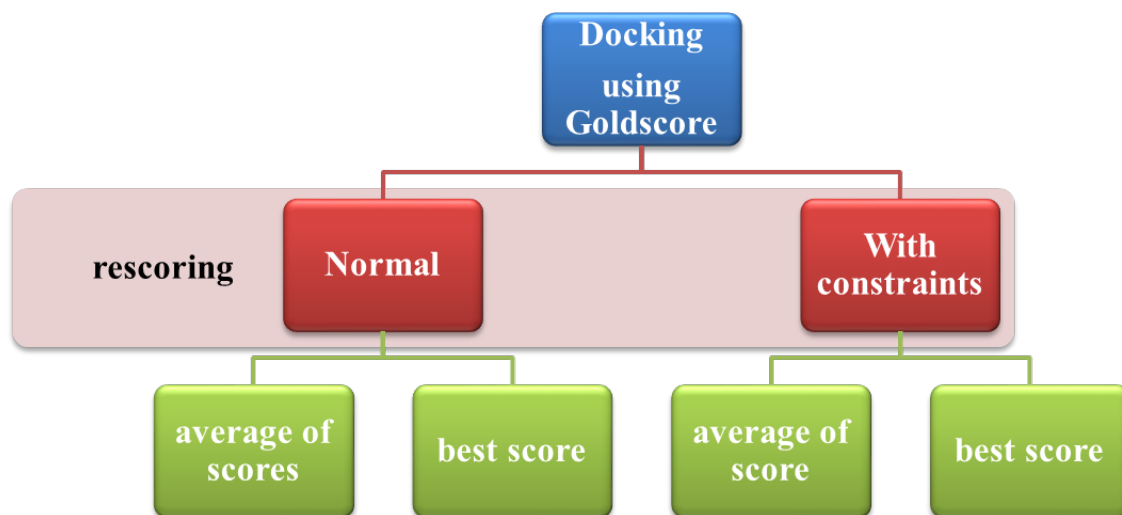


Figure 5 – Diagram summarising the method used to obtain fitness scores for the 1534 compounds of the PPI-Net library.

Ten GA runs were performed and stored for each compound. Despite a prediction of a good fitness score by the scoring function being important, the consistent calculation of relatively high scored poses across several GA runs is also important. Consequently, for each compound, and each scoring function two values were attributed: the best fitness and the average obtained across the ten GA runs.

Finally, and because the compound library is relatively diverse, with compounds from two different suppliers, with molecular weights ranging from 232.21 to 530.51 g.mol<sup>-1</sup>, it was decided to apply a normalisation factor so scores obtained would be more homogeneous and less affected by the size/structure of the molecule. Therefore, *ligand efficiency* (LE) was calculated. This approach was utilised before and is described in Chapter 2 (score/number of heavy atoms).

At this moment the 1534 molecules were organised in four different ranks:

1. LE Rank – normal docking: rank calculated based on the ligand efficiency values obtained for the best score
2. Average LE Rank – normal docking: rank calculated based on the ligand efficiency values obtained for the average score
3. LE Rank – biased docking: rank calculated based on the ligand efficiency values obtained for the best score
4. Average LE Rank – biased docking: rank calculated based on the ligand efficiency values obtained for the average score

It is not clear which rank will give more meaningful discrimination between active and inactive compounds. Ideally, all of them should be incorporated into the analysis. A method used frequently in this project to rank compounds through a balanced selection of different parameters, *Pareto ranking*, was again used. In addition to rank higher the best molecules in each previously described rank, aqueous solubility was again introduced as part of the balanced selection.

At this stage compounds were ranked based on docking predicted scores and aqueous solubility. The next step will be selecting the 96-well plate with probability of having more active hits. For this selection two considerations were taken into account. The first is obviously related with the Pareto ranking, the plate with more compounds in the top ranked positions should, in principle, be considered for testing. A box plot was created based on the values obtained from Pareto ranking *per* plate barcode. The results are in Figure 6 and show plates 2000604-607 occupy not only in a lower position in the plot, indicating small values for the Pareto ranking and therefore higher hierarchy in the rank, but also the boxes are smaller indicating a smaller dispersion of values.

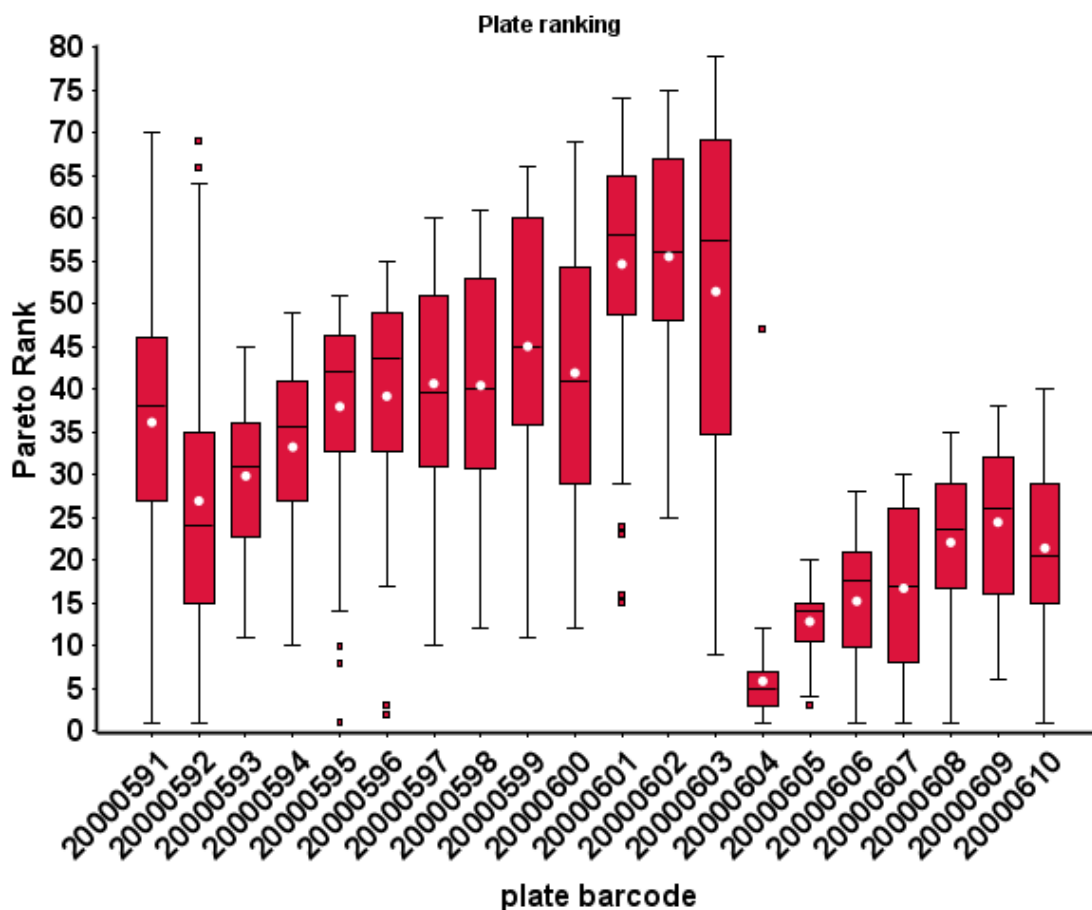


Figure 6 – Box plot based on the Pareto ranking obtained for the 1534 compounds of PPI-Net library. The ends of the whiskers represent the 95<sup>th</sup> and 5<sup>th</sup> percentiles, the ends of the boxes represent the 75<sup>th</sup> and 25<sup>th</sup> percentiles, the horizontal line represents the median and the white circle represents the mean value. Outliers are shown as red squares.

The second is related with chemical diversity, as when performing an initial screening a diverse set of molecules with different properties is desirable as it improves the chances of not only finding hits but also gives better possibilities for hit optimisation. To evaluate chemical diversity DataWarrior<sup>21</sup> was used. DataWarrior is a multi-purpose data analysis tool where data can be filtered on structural motives, views are chemistry aware, molecule properties can be predicted from chemical structures, and specialized cheminformatics methods explore the relationship between chemical structure and measured properties. Compounds belonging to plates (20000)604-607 (for clarity the prefix 20000 from each barcode will be omitted from now on) were grouped according with their structure similarity, based on SkelSpheres descriptor, described as the most accurate descriptor for calculating similarities of chemical graphs in DataWarrior. The map obtained is in Figure 7, where plate 604 seems to be

clustered four big clusters of chemical structure similarity, meaning the compounds within this plate are not very diverse using the SkelSpheres descriptor. Solubility analysis was also incorporated to help plate selection – see legend in Figure 7. Plate 604 solubility is mostly between green (logS around -4 and -5) and yellow (logS around -6). Plate 605 is well dispersed with fewer connections between compounds belonging to the same plate. Solubility values are now better, varying between cyan (logS ~ -3) and yellow (logS ~ -6). Compounds in the plate 606 show good diversity as well. Solubility wise, it looks this plate contain highly soluble compounds. Finally, plate 607 also shows good dispersion. However, few small clusters of structurally related compounds can be observed. Overall, these compounds seem to have good solubility values.

A first analysis would possibly dictate that plates 605, 606 and 607 would probably be the most suitable for *in vitro* screening, because they have better solubility values. However, in Figure 6 considerations about aqueous solubility were already included in the analysis. It was therefore decided to make a balance between predicted affinity and solubility. Plate 605 is clearly the plate with higher ranked compounds based on binding affinity prediction and solubility. In terms of the plate with compounds predicted to bind tighter to EB1c, plate 604 is clearly the front-runner. Finally, it was decided to select these two plates 604 and 605 for Octet® RED96 System screening. It is worth mentioning that four compounds included in these plates are found in the final compound selection obtained from QSAR approach, namely 1030, 1057 and 1071 from plate 604 and 1081 from plate 605.

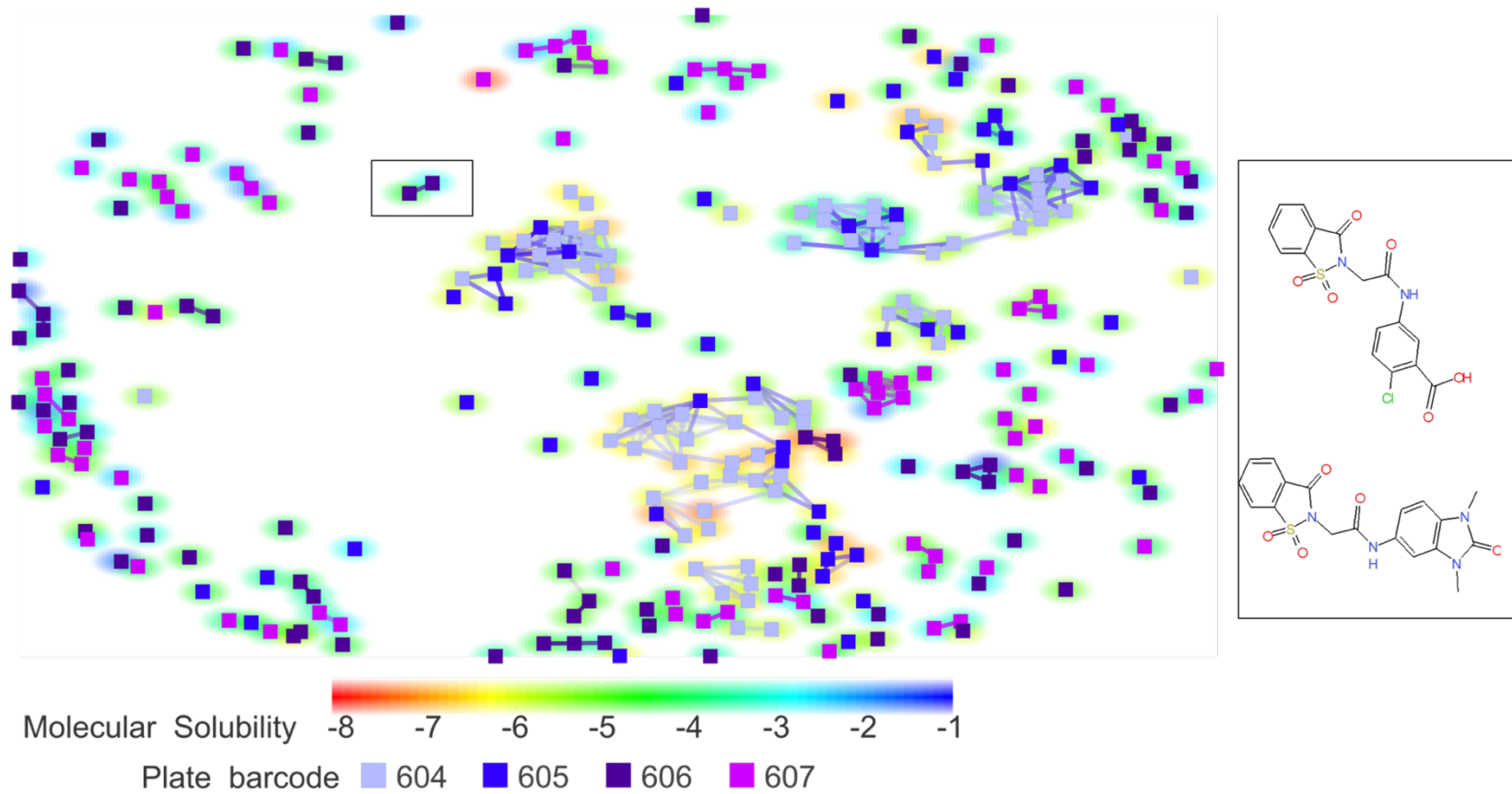


Figure 7 – Dispersion map based for the four 96-well plates to be analysed. Each square represents a different compound, coloured by plate. Connection lines are for neighbour compounds, *i.e.* compounds with similar structures. Background colour is related with the molecular solubility (logS). As an example, the two compounds on the right hand side corresponds to the two compounds from plate 606 selected in the map. These compounds have a SkelSpheres similarity value of 0.83. The figure was produced using DataWarrior<sup>21</sup>.

#### A.4.7 Octet® RED96 System screening

From the PPI-Net library 135 compounds, corresponding to plates 604 and 605 were tested using the Octet® RED96 System. This method is based on attachment of the target protein to a surface support (biosensor) through biotinylation reaction and then dipping the chip (biosensor) in ligand solution. The measurements provide kinetic parameter parameters such as  $K_a$  and  $K_d$ .

This technique normally relies on biosensors that can be re-used.<sup>21</sup> Upon binding of a ligand to a protein one should observe two distinct processes: the association and the dissociation. Figure 8 shows an ideal situation where the association phase can be observed up to 90 seconds followed by dissociation. Aspects related with this method will not be further detailed as this was an exploratory project which aim was to try to find a HTS method that could be applied for protein-protein interactions, as an alternative to biophysical techniques used (NMR and ITC).

From the previous section results, it was decided to test all the compounds present in plates 604 and 605, yielding a total of 135 compounds. A 384-well plate was used to screen the selected compounds, where each well containing compound was accompanied with a blank well (without ligand) with dissociation buffer. Ligand concentration in each well was 500  $\mu\text{M}$ .

Nineteen compounds were selected as possible ligands for EB1c $\Delta$ 8, 11 from plate 604 and eight from plate 605 – Table 6. The average molecular weight for the selected compounds is  $435 \pm 16 \text{ g}\cdot\text{mol}^{-1}$ . In terms of aqueous solubility these are more diverse, with logS values from -4.4 to -7.0, indicating that some of these compounds may not be suitable for testing using NMR or ITC due to their poor aqueous solubility. These compounds show a profile consistent with the association/dissociation pattern expected for a favourable interaction with the protein – Figure 9. None of these compounds correspond to the four compounds common to the QSAR final dataset and the tested compounds.



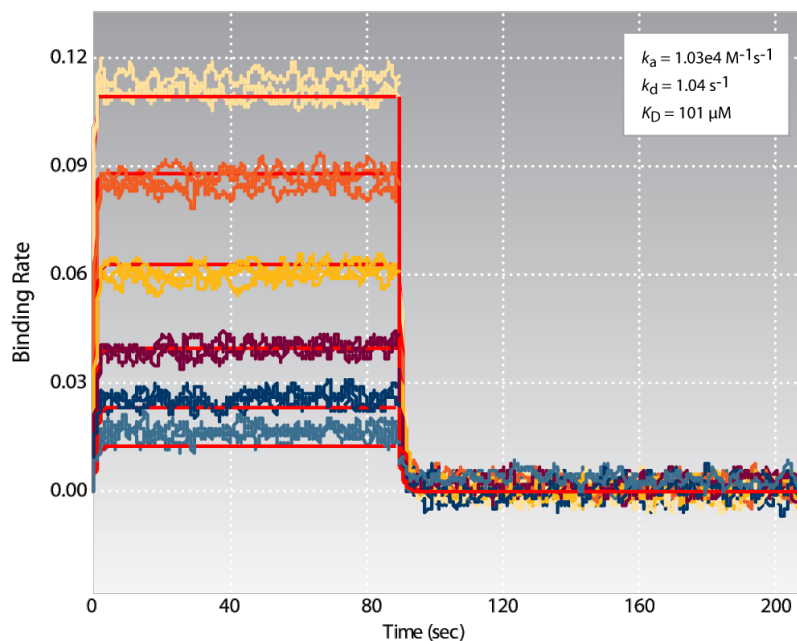


Figure 8 – Example of association/dissociation data normally obtained from Octet® RED96 System. The image was obtained from the supplier’s website <http://www.fortebio.com/octet-RED96.html>.

Table 6 – Compounds selected as active after Octet® RED96 System screening.

Octet Well	Plate barcode	PPI-Net plate well	Vendor ID	PPI-Net ID
A9	20000604	A10	BDH34136166	1064
C6	20000604	B7	BDH32350590	1041
C7	20000604	B8	BDG34020588	1049
C9	20000604	B10	LAS34137204	1065
E1	20000604	C2	LAS34122636	1002
E2	20000604	C3	LAS34124173	1010
I8	20000604	E9	BDG34020575	1060
K7	20000604	F8	BDG34020582	1053
K8	20000604	F9	BDH34136185	1061
O7	20000604	H8	BDH32352181	1055
O8	20000604	H9	BDG34137236	1063
A12	20000605	A3	BDG340205579	1088
C13	20000605	B4	BDG34127776	1105
C14	20000605	B5	BDG34018779	1097
E13	20000605	C4	BDG340020587	1098
E17	20000605	C8	BDG34125650	1130
I15	20000605	E6	BDG34136119	1116
K17	20000605	F8	LAS34135965	1133
O14	20000605	H5	BDG34129080	1111

The remaining compounds showed either no response or the curve obtained was not consistent with association/dissociation pattern. Some showed only association and no dissociation and for some the association/dissociation curve gave opposite responses to the ones expected. A possible explanation may be non-specific binding to the surface.

Unfortunately, many of the association/dissociation plots were inconclusive since some biosensors, when reused, were contaminated by ligands that failed to dissociate at the wash stage, thus affecting the subsequently analysed ligands using the same biosensor. This problem shows two limitations of this technique for HTS. The first is related with the speed and automation of the technique. In order to be a good HTS method it should be fast and automated.<sup>21</sup> Replacing the contaminated sensors would require stopping the analysis which would need both time and supervision. Secondly, the reusability of the biosensors was one of the manufacturer's selling point and from a HTS point of view one wants not only a reliable technique that can screen many compounds in a short period of time but also inexpensive enough to be a HTS technique. Buying a biosensor for each compound is expensive not only in terms of consumables but in terms of protein used.

Another difficulty found when testing Octet® RED96 System is the low limit for compound concentration. This may be a complication when doing a first HTS in order to find active compounds. One does not expect, at least at a very initial stage to have highly active compounds. Alternatively, and because this technique allows for the calculation of parameters such as  $K_a$ ,  $K_d$ , this technique seems to be suitable for hit validation, and in that case as a complementary method to a HTS screening. However, when testing the active molecules found to bind to EB1cΔ8 by solution NMR the results showed these compounds did not interact with EB1. A possible reason for that is the fact that high concentrations of these molecules are needed to have an interaction since these are weak binders. Ideally, one would like select some of these compounds, both active and inactive, do the opposite test, screening them using solution NMR. Additionally, testing molecules 1030, 1057, 1071 and 1081, by NMR could be another interesting, since these are common to both virtual screening approaches used to screen this library.

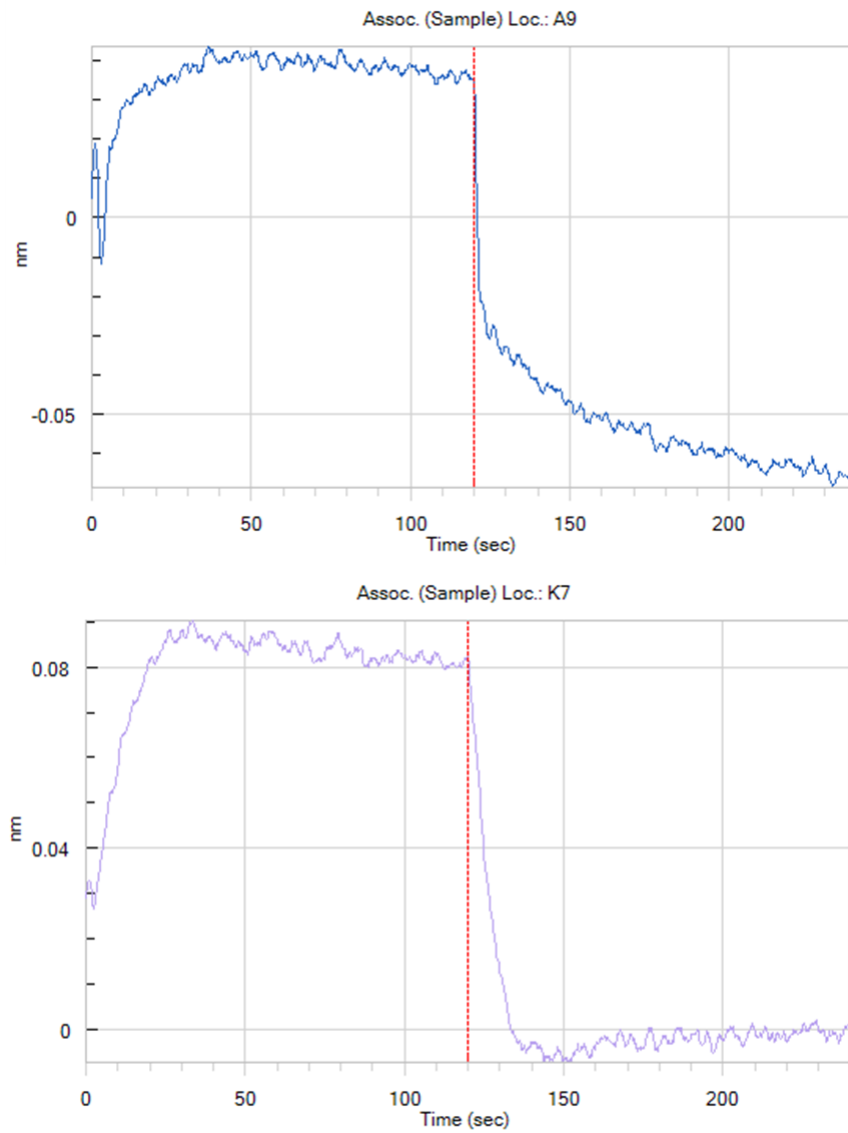


Figure 9 – Data obtained from Octet® RED96 System screening for compounds 1064 (top panel) and 1053 (bottom panel) as examples of positive hits for EB1c binding.