Conformational Control by Intramolecular Hydrogen Bonding

James W. Luccarelli

A dissertation presented in partial fulfilment of the requirements for the award of the degree of

Doctor of Philosophy

of the

University of Oxford

Chemistry Research Laboratory
12 Mansfield Road
Oxford OX1 3TA

Oriel College
Oriel Square
Oxford OX1 4EW

Trinity Term 2013

Authorship Statement

This thesis describes work conducted at the National Institutes of Health, Bethesda MD USA; and the Chemistry Research Laboratory, University of Oxford. It is my own work, except for the following sections:

Chapter 1: The section on literature inhibitors of SH3 domain binding (Ch 1.4) was written in conjunction with Sam Thompson (postdoctoral fellow, Hamilton group, Dept of Chemistry, Oxford). Portions of this chapter were previously published:

Luccarelli, J., Thompson, S. & Hamilton, A.D. "SH3 Domains as Drug Targets" *Protein-Protein Interactions in Drug Discovery*, Ed. A. Doemling. Wiley-VCH, 2013. 101-128.

Chapter 2: Grb2 SH3C protein was expressed and purified by Philip Simister (postdoctoral fellow, Feller group, Dept of Oncology, Oxford). Dr. Simister also designed the competition binding assay and performed the control experiments validating the assay. Portions of this chapter were previously published, with Dr. Simister and I sharing joint first-authorship:

Simister, P.C.,[†] Luccarelli, J.,[†] Thompson, S., Appella, D.H., Feller, S.M., & Hamilton, A.D. Novel inhibitors of Grb2 SH3C Domain Identified by a Virtual Screen. *Bioorg. Med.Chem.*, **21**, 4027-4033 (2013).

Chapter 3 & Chapter 4: Grb2 SH3C protein was expressed and purified by Philip Simister.

James Luccarelli

Table of Contents

Conformational Control by Intramolecular Hydrogen Bonding

James Luccarelli, Oriel College
DPhil, Trinity 2013

Abstract

Hydrogen bonds are directional, non-covalent interactions between hydrogen and electronegative atoms. Although generally weak, these interactions are critical to the stability of many biological systems including proteins and DNA. This dissertation explores small molecules in which an intramolecular hydrogen bond is the key determinant of conformation.

Chapter 1 introduces the protein Grb2 SH3C, details its role in cancer signalling, and delineates the idea of peptidomimetics—small molecules which are functionalized to mimic the structure of a peptide and disrupt protein-protein interactions.

Chapter 2 describes a virtual screen for binders to Grb2 SH3C. From a library of 6.3 million compounds, 34 were tested *in vitro* and two found to bind to the protein in two orthogonal assays.

Chapter 3 describes mimics of the polyproline II helix using a benzoylurea scaffold. A small library of these compounds was synthesized and tested for binding to Grb2 SH3C using SPR, a competition assay, and NMR.

Chapter 4 describes attempts to mimic a 3_{10} helix using benzamide-based peptidomimetics. The synthesis and *in vitro* evaluation of these molecules as ligands of Grb2 SH3C is described.

Chapter 5 uses quantum chemical calculations to assess the energies of a series of molecular switches. These calculations benchmark a range of modern density functional theory calculations, and attempt to quantify the accuracy of these methods for a large, flexible system. The role of solvation, entropy, geometry, and torsional angles are assessed in accurately calculating the energies of the critical hydrogen bonds.

Acknowledgements

Although we didn't know it at the time, this thesis was born in 2008 when a chance meeting in Bill Jorgensen's office introduced me to Prof. Hamilton. At the time I had only been a chemistry major for 6 months, and was planning to start medical school after graduation, but that meeting and those that followed led me to this journey "across the pond." I'm extremely grateful to Andy for taking a risk on a student who had never even run a column before. I've learned a tremendous amount about synthesis and chemical biology, and have been privileged to have you as a mentor.

Since coming to Oxford, I've had the pleasure of working very closely with Dr. Rob Paton on computational organic chemistry. His insights into contemporary methods, life as a new faculty member, and 90s rap have been invaluable additions to my Oxford education. Additionally, my NIH supervisor Dr. Dan Appella helped me to navigate the world of government chemistry. I know that more time in your lab would have been a fantastic experience, and I'm sorry I didn't get the chance to work more closely with you during my DPhil.

My research was supported by the Marshall Aid Commemoration Commission, NIH, Oriel College, and the University of Oxford. Each has contributed financially and in invaluable non-financial ways, and I appreciate their investment in my education. Also thanks to Harvard Medical School for giving me this time to devote to research without having to worry about where I'll be next.

The analytical staff in the CRL, particularly Dr. Tim Claridge, have helped me to make the most of Oxford's fantastic instrumentation. My colleagues during my DPhil have made my time here so much fun: Dr. Philip Simister has been an ideal collaborator who has added great insight to the project and always made sure I wasn't concluding more than the data showed. The Happy Hamiltons have been a constant joy to work with: Dr. Sam Thompson

got me started on the project and began my cricket, ale, and (bow-tie) solvent screen education. Dr. Hannah Lingard has been a fantastic experimental checker and even better friend, and has introduced me to the life-changing concept of “the drop.” prof. marc adler stressed the importance of checking ether for freshness and not always listening to sam. Dr. Hayden Peacock has been my rock in biology and gliding.

My fellow DPhil students are welcome additions to the lab: Madura Jayatunga has shaped up to be an adequate future Senior DPhil and talented wordsmith. Jonathan Ross’s infectious love of amide coupling reactions, rowing, and the education of Oxford 1st years has been a boon to the lab. Tom Han kept me on my toes during his rotation project, and his board game nights have brought the department closer together. Part II-turned-*pseudo*-biochemist Daniel Yin spends more of his free time on chemistry than anyone I’ve ever met; I look forward to reading his pioneering work in JLC. The six part IIs and rotation student Cyrille Thinnes also contributed in pretty much every way to the life of the group, and I wish them luck as they move onwards and upwards.

My parents and brothers have been wonderfully supportive as I’ve disappeared for three years. It will be a bit difficult adjusting to life post-Hogwarts, but I’m looking forward to coming home.

Finally, and most importantly, to Ingrid, who has done more than anyone else to shape my time in Oxford. Thank you for being wonderful.

Glossary of Abbreviations and Acronyms

Å – Angstrom

Ac – acetyl (CH_3CO-)

Arg – arginine

ATR – attenuated total reflection

Bn – benzyl

Boc – *tert*-butoxycarbonyl

B3LYP – Becke exchange with Lee-Yang-Parr correlation (functional)

C – Celsius

Cal – calorie(s)

CAM-B3LYP – Coulomb-attenuated B3LYP (functional)

CBS – complete basis set

CBS-QB3 – composite complete basis set (method), using B3LYP geometries and frequencies

cc-pVQZ – correlation-consistent polarized valence quadruple zeta (basis set)

cm – centimetre(s)

COSY – correlation spectroscopy

CPCM – conductor-like polarisable continuum model

cProMe – cyclopropylmethyl

δ – chemical shift

Da – Daltons

DCM – dichloromethane

DFT – density functional theory

DIPEA – ethyldiisopropylamine

DMAP – 4-dimethylamninopyridine

DMF – *N*,*N*-dimethylformamide

DMS – dimethylsulfide

DMSO – dimethylsulfoxide

DPA – diphenyl acetylene

EDC – 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

EI – electron impact

ESI – electrospray ionization

Et – ethyl

Ether – diethyl ether

FI – field ionization

g – gram(s)

ΔG – Gibbs free energy

Gab2 – Grb2-associated binder

Grb2 – growth factor receptor-bound protein 2

GGA – generalized gradient approximation

HF – Hartree-Fock

HMBC – heteronuclear multiple bond coherence

HPLC – high-performance liquid chromatography

HRMS – high-resolution mass spectrometry

HSQC – heteronuclear single quantum coherence

Hz – Hertz

i – *iso*

IR – infrared

IUPAC – International Union of Pure and Applied Chemistry

Lys – lysine

Me – methyl

M06-2X – high nonlocality Minnesota 06 (functional)

MUE – mean unsigned error

NHS – *N*-hydroxysuccinimide

NMR – nuclear magnetic resonance

PDB – protein data bank

Petrol – petroleum ether boiling in the range of 30 – 40 ºC

Pd/C – palladium on activated charcoal

PI – predictive index

PPI – protein-protein interaction

PPII – polyproline II helix

Pro – proline

Quant. – quantitative

r^2 – linear coefficient of determination

RMSD – root mean square deviation

SAR – structure-activity relationships

SH2 – Src homology 2

SH3 – Src homology 3

S_N2 – substitution nucleophilic bimolecular

S_NAr – substitution nucleophilic aromatic

SPR – surface plasmon resonance

TBAF – tetra-*n*-butylammonium fluoride

TBS – *tert*-butyldimethylsilyl

tert – tertiary

TFA – 2,2,2-trifluoroacetic acid

THF – tetrahydrofuran

TLC – thin-layer chromatography

υ_{max} – maximum wavenumber

ωb97X-D – long-range corrected density functional incorporating dispersion corrections

x – any amino acid

1. Introduction: SH3 domains as drug targets

1.1. Protein-protein interactions

Virtually all cellular processes involve interactions between multiple proteins. Among these are protein-protein interactions (PPIs), which allow for dynamic information exchange and the precise control of cellular function beyond the genetic coding of proteins. Indeed, the quantity and type of PPIs is hypothesized to be more important to an organism's complexity than the actual size of the genome.[1] This was made abundantly clear following the sequencing of the human genome, when the total number of human genes (ca. 20,000) was found to roughly equal that of the flatworm *C. elegans*.[2,3] The total number of interactions in humans, however, is much greater, with an estimated 650,000,[4] compared to ca. 120,000 for *C. elegans*.[5]

Signal transduction is a particular area of cellular function that relies on PPIs.[6] In multicellular organisms, extracellular signalling molecules are key determinants of a cell's state. In general, these molecules are detected by protein receptors on the cell surface, which change conformation in response to the signal. This conformational change is detected by intracellular proteins, which, *via* a series of interactions with other proteins, ultimately cause propagation of the signal.[7] Often these interactions occur at the level of domains—stable units of protein structure that fold autonomously.[8] Domains are often specific for particular binding targets, and several can be combined to form larger proteins, allowing multidomain proteins to bind multiple targets.[9] This versatility is enhanced by specialized adapter proteins, which like plug adapters for electrical appliances fit between proteins with different signalling domains to allow them to interact.[10] Critically, several protein interaction domains are conserved throughout life, indicating that the interactions mediated by these domains are similarly conserved.[11]

1.2. SH3 domains

The Src homology 3 (SH3) domain is one such conserved fold critical to cell signalling.[12–18] Since the discovery of the first of these domains in a viral Src oncogene[19] and phopholipase-C[20] in 1988, researchers have identified more than 300 in the human genome.[21] SH3 domains are found across biology: in addition to sarcoma viruses and man, the Pfam database[22] lists 79 proteins in *C. elegans*, 175 in *Drosophola*, and 551 in *M. musculus* which contain at least one SH3 domain. SH3-like domains are also found in prokaryotes, including bacteria[23] and archaea.[24] In humans, SH3 domains are found in a wide range of structures, including tyrosine kinases (*e.g.* c-Src, Abl, and Fyn), adaptor proteins (*e.g.* Grp2, Grb2, Nck, and Crkl), GTPase-activators (e.g. Graf), lipase (PLC-γ1), and a structural protein (spectrin), among others.

1.2.1. Structure

Three dimensional structures of spectrin[25] and Src[26] SH3 domains were solved via X-ray crystallography and NMR, respectively, in 1992. Since then nearly 500 structures of SH3 domains have been deposited in the PDB. SH3 domains contain approximately 50-70 amino acids, arranged in a β-sandwich consisting of five β-strands (Figure 1-1). These strands form two anti-parallel β-sheets of three strands each: the first comprising strands **a**, **e**, and the first half of **b**, and the second strands **c**, **d**, and the second half of b. The β-sandwich resulting from the interaction of these sheets forms the hydrophobic core of the domain.

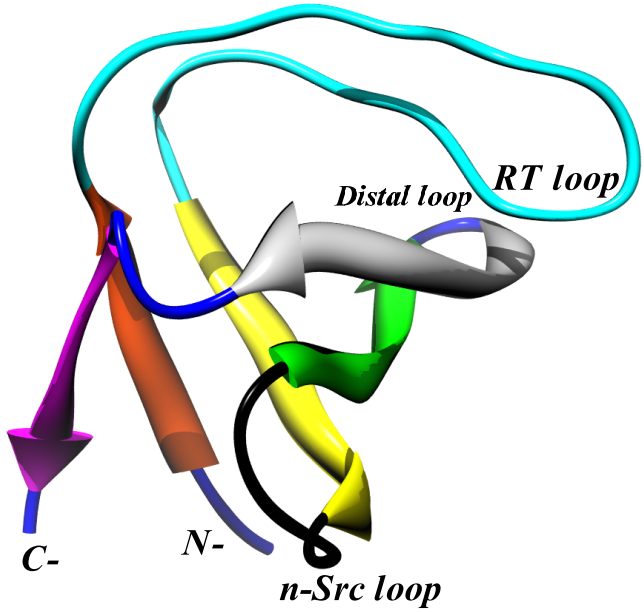


Figure 1-1: X-ray structure of the SH3 domain of Src (PDB ID 1SHG)[25] showing strand **a** (red), RT loop (cyan), kinked **b** strand (yellow), n-Src loop (black), **c** strand (green), **d** strand (grey), and **e** strand (purple). The remaining loops are coloured dark blue. Molecular graphics images were produced using the UCSF Chimera package.[27]

The β-strands are connected by three loops. Strands **a** and **b** are separated by a long RT loop, which generally contains several charged residues and forms an irregular antiparallel structure. A neuronal Src (n-Src) insertion loop connects strands **b** and **c**, while a short distal loop (which lies opposite the other two loops) connects strands **c**, and **d**. The active site of the protein is the groove defined by the RT and n-Src loops. Strands **d** and **e** are often connected by a short stretch of 3_{10} helix, bringing the *N*- and *C*- termini close in space on the face of the domain opposite the binding site.

1.2.2. Variability

The thousands of SH3 domains identified across biology show only moderate sequence conservation. The pair-wise identity among a large series of SH3 domains, comparing each sequence in the database to every other one, was on average 27 %, with most in the range of 15 % to 45 % (Figure 1-2).[28] Much of the variance in sequence length occurs within the n-Src and distal loops, with general conservation of the length of the RT loop. A series of mutagenesis experiments on the core of the Fyn SH3 domain showed that mutation of the

conserved hydrophobic residues resulted in greatly decreased protein stability, indicated by reduced denaturation temperatures.[29] Particularly important were the residues flanking the RT loop, which are conserved as Asp and Leu, respectively in 75% of domains.[28] Mutation of these residues reduced the denaturation temperature of Fyn SH3 from 80.1 ºC to as low as 36.3 ºC for L18A, indicating that they are critical to the stable fold of most SH3 domains.[29] While great sequence variability has been observed for SH3 domains, relatively little structural variability has been found. Overlays of multiple structures have been conducted, with average root mean square deviations (RMSD) of less than 2 Å for the residues within the β strands.[28]

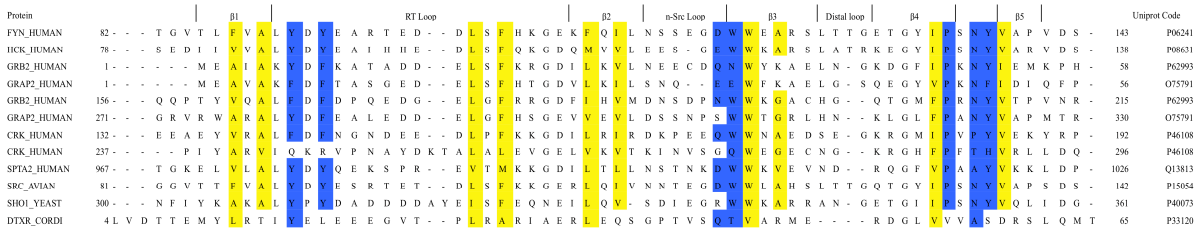


Figure 1-2: Sequence alignment of a series of SH3 domains from humans and other species. Conserved positions in the hydrophobic core are shaded yellow, while conserved ligand-binding residues are highlighted in blue. Uniprot accession codes for each protein are listed after the sequence. Sequence alignment was performed using the ClustalW algorithm.[30] The secondary structure boundaries listed above the sequence are those for the *C*-terminal SH3 domain of Grb2.

1.2.3. Binding motifs

NMR studies confirmed the SH3 binding site to be the aromatic-rich region between the RT and n-Src loops.[26] Further work identified the canonical SH3 domain binder as a proline-rich peptide.[31] Several structural features of proline make it well-adapted as a cellular signalling motif.[32] It is unique among the natural amino acids in containing a cyclic secondary amine as its sidechain. This sidechain is conformationally constrained, giving rise to a rigid structure of fixed dihedral angle that is capable only of accepting, not donating, hydrogen bonds. Moreover, proline *C*-terminal amide is significantly more likely than any other amino acid to adopt the *cis*-amide conformation, as its barrier to *cis*/*trans* isomerization is lower.[33] The overall effect is to allow many proline-rich sequences to form

a unique, left-handed polyproline II helix (Figure 1-3).[34] PPII helices have a three-fold symmetry formed by three amino acids per turn, with a distance between the *i* and *i*+3 residues of approximately 9 Å.[35] Sidechains are projected roughly perpendicular to the helix axis. Hydrogen bonding is primarily to solvent,[36] and amino acids with limited steric bulk that do not shield the backbone from solvent are relatively favourable for PPII helix formation.[37]

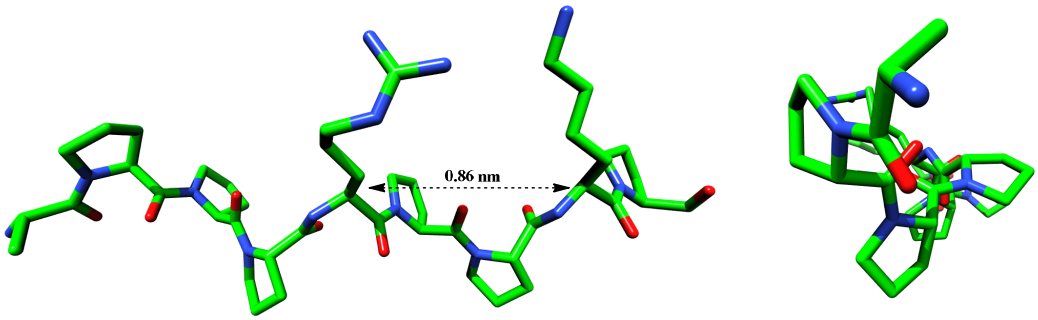


Figure 1-3: Perpendicular (left) and axial (right) view of the PPII helix of Gab2a (APPPRPPKP), which binds to the *C*-terminal SH3 domain of Grb2 (PDB ID 2W0Z).[38] The extended helix exposes the backbone of each residue to solvent, and has three-fold symmetry about its axis.

Using a biased combinatorial peptide library, Chen *et al.* identified two different classes of sequences that bind to SH3 domains.[37] The consensus sequence for the class I ligands is [R/K]xXPxXP, while the class II sequence is PxXPx[R/K], where x signifies any amino acid and X signifies any non-glycine hydrophobic amino acid. Due to the symmetry of the PPII helix, a 180º rotation perpendicular to the helix results in a very similar structure: only the *N*- and *C*-termini, and the directionality of the XP dimer changes with this inversion. Thus, the invariant proline residues of class I and class II ligands occupy different sides of the PPII helix depending on the orientation of the ligand. The binding affinities of the peptides identified ranged from 7 μM – 30 μM.[39,40]

Structural evidence has been found for the binding of both class I[41] and class II[42] orientations. Based on these data, the Schreiber group proposed a general model for SH3 domain binding relying on a specific hydrogen bond from the basic residue in the ligand and

extensive hydrophobic contact between proline residues and the SH3 binding pocket (Figure 1-4).[43] Under this model, the SH3 domain has distinct binding pockets defined by conserved hydrophobic residues. The two essential prolines from the ligand bind into these pockets, making extensive hydrophobic contacts which provide much of the favourable energy of the interaction. While these hydrophobic pockets are not very deep or distinct, they show remarkable selectivity for proline residues. Detailed investigations by Nguyen *et al*. have shown that this recognition preference is specifically for *N*-substituted amide residues.[44] Proline is thus required for binding as it is the only common naturally-occurring *N*-substituted amino acid, and not due to a specific steric preference for the proline sidechain.

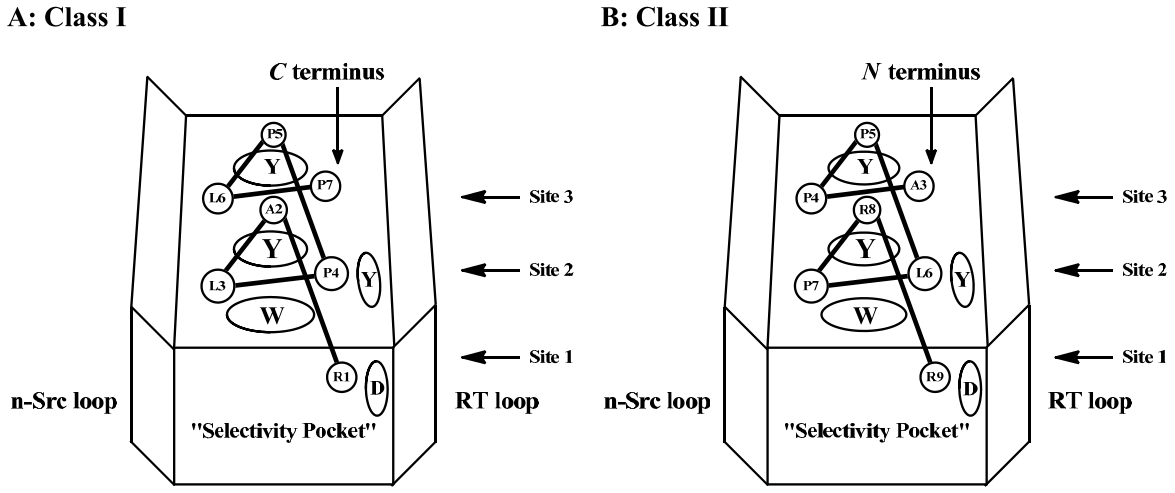


Figure 1-4: Schematic showing the binding positions for a class I (A) and class II (B) ligands bound to c-Src. Ligand residues are indicated by circles, and amide bonds as sticks. Conserved SH3 domain residues are shown as ovals. The class I ligand is RLP2 (RAL*P*PL*P*RY), while the class II ligand is PLR1 (AFA*P*PL*P*RR). The critical proline residues are italicized. In each case, a salt bridge between a conserved Arg residue in the ligand and an SH3 domain Glu determines the orientation of the ligand (site 1). The essential Pro residues bind in hydrophobic pockets (sites 2 and 3) formed by conserved aromatic residues in the SH3 domain.

While the majority of work on SH3 domains has focused on the classical class I and class II domains, certain atypical binding motifs have been recognized in specific proteins. For instance, the Grb2-related adapter protein 2 (Grap2)[45] binds to the signalling protein SLP-76 through an RxxK motif.[46] NMR[47] and crystallographic[48] analysis of the complex revealed the specific molecular interactions of the site: the cationic peptide arrays the

critical residues in a 3_{10} helix (*vide infra* Chapter 4.1), as opposed to the typical polyproline helix. This binding displayed sub-μM affinity—more than two orders of magnitude stronger than observed for PxxP ligands—with the mutagenesis of either cationic residue to alanine completely destroying binding affinity.[47,48] In contrast, replacement of proline residues in the sequence had a deleterious effect on binding affinity, but did not abolish binding.

1.2.4. Selectivity

Since SH3 domains are involved in the assembly of multi-protein structures, their selectivity for their targets could greatly affect signal transduction. One possibility is that *in vivo* binding is generally non-selective: that is, rather than existing as definite, linear pathways of **A→B→C**, the reality is a combinatorial series of interactions, with **A** binding to **P**, **Q**, **R**, **S** and **T**, and each of those in turn having other binding partners in addition to **A**.[49] Under this view, should a protein develop a new SH3 domain or binding sequence (e.g. by exon shuffling or gene duplication), it would begin to interact with numerous other proteins, thus subtly altering the equilibrium of the cell. This change would occur without a concurrent evolution of a specific binding partner for the new domain, thereby eliminating the need for all SH3-mediated PPIs to have evolved selectively.[50]

In contrast to this theory, a great deal of intrinsic SH3 domain selectivity has been found in yeast. The response by *S. cerevisiae* to high osmolarity involves the binding of a proline-rich region in Pbs2 to the SH3 domain of Sho1.[51] To probe the selectivity of this interaction, Zarrinpar *et al.* created a library of Sho1 chimeras, in which the SH3 domain of the wild-type protein was replaced with other SH3 domains.[52] Of the dozen metazoan SH3 domains tested, six were able to restore function of the osmotic pathway *in vivo*, and the same six bind Pbs2 in *in vitro* assays. This high degree of promiscuity is consistent with the notion that SH3 domains are not intrinsically selective. On the other hand, when any of the other 26 yeast SH3 domains were inserted into Sho1, none restored function *in vivo* or bound the

peptide *in vitro*.[52] This result was explained with a model of "negative evolution," whereby proteins in an organism evolve not merely to bind a target, but also to do so in a manner orthogonal to other proteins within that organism (Figure 1-5). Thus, while the overall selectivity of yeast SH3 domains to a library of peptides may be low, selectivity for those peptides in the yeast proteome is absolute. Mutating the sequence within the Pbs2 motif and again comparing cross-reactivity for other yeast SH3 domains probed this evolutionary pressure. While some of the mutations actually increased binding affinity for Sho1, all increased cross-reactivity,[52] indicating that evolutionary pressure in this system did not seek merely high binding affinity.

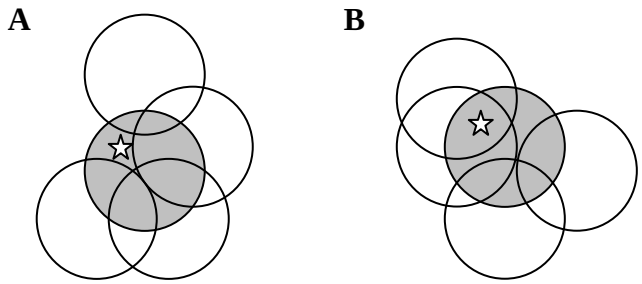


Figure 1-5: Model for domain-mediated selectivity via "negative evolution." (A) recognition profiles within an organism (circles). These profiles are substantially overlapping, but evolutionary selection against cross-reactivity drives each domain to a niche sequence (star, for the grey domain). (B) recognition profiles among different organisms. Once again these profiles are substantially overlapping, but as evolutionary pressure for selectivity has not been applied there is significant cross-reactivity for the starred binding sequence. Figure adapted from Zarrinpar *et al*.[52]

A more recent experiment investigated whether human SH3 domains show similarly-limited cross-reactivity. Seet *et al*. probed the selectivity of the RxxK motif in SLP-76 for 147 different human SH3 domains, and found that only four bound.[53] Of these, the physiologic binding partner Grap2 bound with 8 nM affinity, a factor of 1000 stronger than Grb2 SH3C, which bound with 8 μM affinity. The remaining two binders, STAM1 and STAM2, are endocytic proteins which have not been linked to SLP-76 *in vivo* and may be localized in a different cellular compartment. Furthermore, a series of point mutations of the SLP-76 peptide were found to increase cross reactivity to different SH3 domains while

maintaining high affinity for Grap2. Thus the atypical RxxK recognition domain in SLP-76 appears to have evolved both for high affinity and high selectivity, but despite this still displays some cross-reactivity to other SH3 domains. The authors repeated these experiments with the RxxK motif of Gab1, the PxxPxRxxK motif of HPK1, and a separate class II PxxP ligand in HPK1 and found binding to 15, 12, and 53 SH3 domains, respectively. As only a fraction of human SH3 domains were screened in this study, it is likely that the cross-reactivity of each of these peptides is even higher.[53]

While further work is required to probe the selectivity of human SH3 domains, a multitude of factors are likely exploited within the proteome. For some interactions (e.g. SLP-76 with Grap2), domain-mediated evolution may be sufficient to explain the very high selectivity observed. This is not, however, general for other proteins, even those sharing an atypical binding motif, as other RxxK peptides are less intrinsically selective. The most common PxxP motif is the least selective, and it is quite possible that the majority of SH3 domains which bind this sequence bind many proteins *in vivo*. To a certain extent this may be tempered by spatiotemporal isolation of proteins, but overall SH3-mediated interactions within human cells likely form a complex web much more than simple, linear chains.

1.3. Grb2 SH3C

One SH3 domain-containing adapter protein that has been extensively studied for its role in human disease is growth factor receptor-bound protein 2 (Grb2). This ubiquitously-expressed, 25 kDa signalling protein consists of three protein recognition domains: a central SH2 domain (which binds phosporylated Tyr)[54] flanked by two SH3 domains.[55] Grb2 lacks enzymatic activity but is essential in a range of signalling pathways, so much so that a targeted gene disruption of Grb2 is lethal to mice at an early embryonic stage.[56] Frequently Grb2 acts as intermediary between receptor tyrosine kinases and intracellular processes.[57,58] Mounting evidence points to a strong role for Grb2 in developing malignancies.[59] The

protein is encoded by the *grb2* gene on human chromosome 17, which is duplicated in some solid tumours and lukemias.[60] Grb2 is also overexpressed in a range of human breast cancer cell lines[61] and tissue samples.[62]

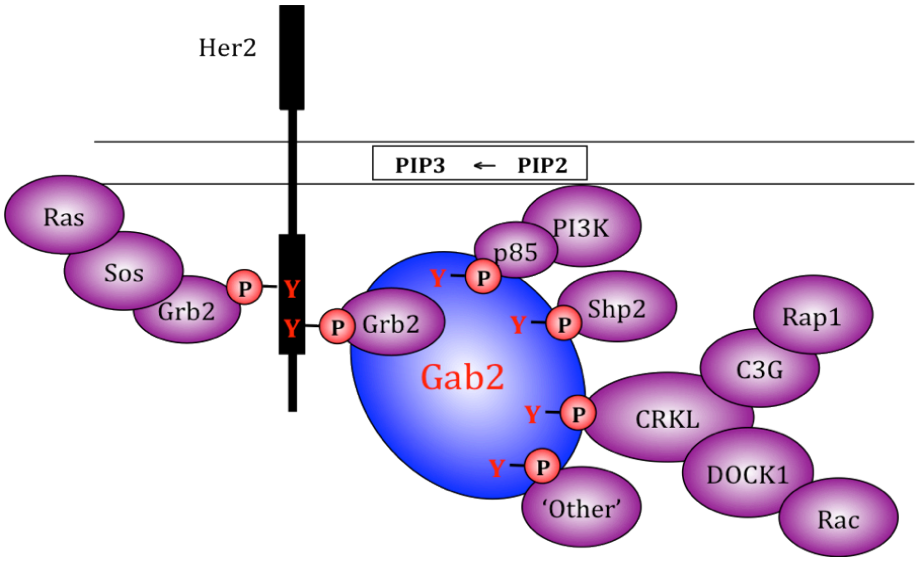


Figure 1-6 The role of Grb2 in the Her2 signalling pathway. Grb2 binds to phosphotyrosine residues on Her2 *via* its SH2 domain, then activates a range of downstream pathways *via* Sos or Gab2.

In breast cancer, Grb2 lies downstream of the EGF family protein Her2 (ErbB2, Neu). Grb2 docks to specific pTyr residues in Her2 *via* its SH2 domain,[63] then recruits Sos through both SH3 domains,[64] or the Gab family of large multi-site docking proteins *via* its *C*-terminal SH3 domain (SH3C).[65,66] This causes downstream activation of a range of pathways, including those involving Ras, PI3K, and Shp2 (Figure 1-6).

Figure 1-7 Kaplan-Meier curve comparing the disease-specific survival of cancer patients with and without Her2 amplification. In this retrospective study of tumour samples collected prior to the discovery of targeted Her2 therapies, median survival in the Her2 amplified group was 2.6 years, compared to 11.7 years in the non-amplified population. Figure from Ainsworth *et al.*[67]

Overexpression of Her2 occurs in ca. 20 % of breast cancers, and is associated with poor survival prognosis (Figure 1-7).[68] Four therapies targeted towards Her2 have been approved for the clinic. Treatment with Trastuzumab, a monoclonal antibody targeting the extracellular domain of Her2, improves survival for patients with these cancers. While large clinical trials have revealed benefits from trastuzumab therapy in metastatic cancer[69] and as an adjuvant therapy,[70] patients rapidly develop resistance.[71] Lapatinib, a small-molecule inhibitor of the Her2 tyrosine kinase domain, has also shown clinical benefits in patients with metastatic Her2-positive cancers, but most patients did not respond to therapy and disease progression continued after a short time.[72] In 2012 the anti-Her2 antibody pertuzumab was approved by the FDA in combination therapy with trastuzumab.[73] The antibody targets the dimerization of Her2, increasing the median progression-free survival time for metastatic disease from 12- to 18-months. Most recently, the antibody-drug conjugate trastuzumab emtansine received FDA approval in February 2013. This drug covalently attaches the cytotoxic agent mertansine to Lys residues on tratsumab. Each antibody binds zero to eight mertansine molecules and delivers them selectively to Her2-expressing cells.[74,75] Phase III clinical trials indicate that this drug improves survival from metastatic breast cancer by approximately six months vs. lapatinib and cytotoxic chemotherapy, but median survival was still only 31 months.[76] Thus while this trial strongly suggests a role for combination therapies in treating Her2-positive cancers, there remains a significant need for novel pharmaceuticals.

One approach is to target the Her2 pathway downstream of the receptor.[77] Cell-based studies have shown that down-regulation of Grb2 inhibits growth in tumours expressing high levels of Her2.[78] Likewise, knockdown of Gab2 reduces proliferation of Her2-overexpressing cells.[79] Further *in vivo* studies in mice have identified a peptide inhibitor of Grb2 SH3 domains which may slightly reduce cancer progression in a xenograft model.[80]

While inhibitors have been developed for the SH2 and SH3N domains of Grb2,[59] small-molecule inhibitors of the SH3C domain have not been reported.

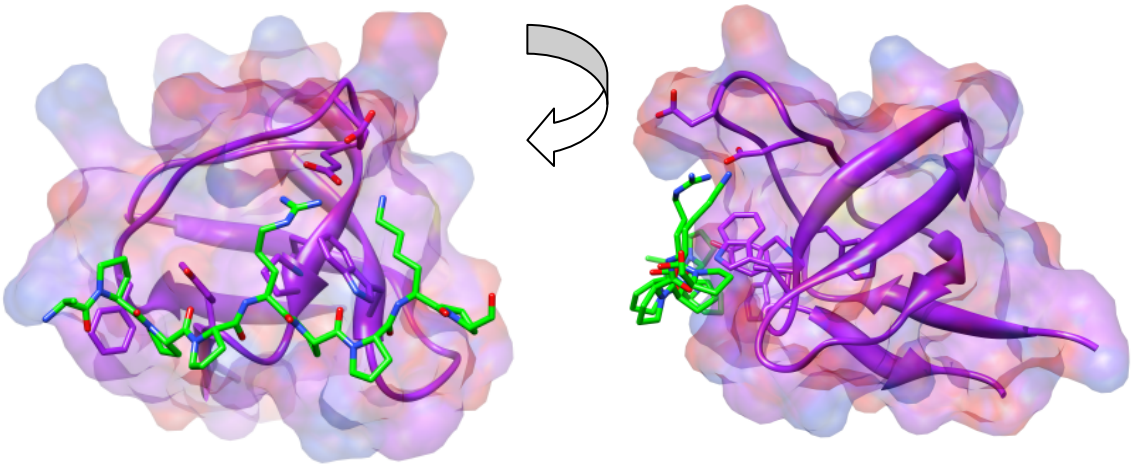


Figure 1-8 The X-ray crystal structure of Gab2a peptide (green) bound to Grb2 SH3C (purple). PDB ID 2W0Z. The interaction is stabilized by hydrogen bonds from the conserved Arg and Lys of the peptide as well as extensive hydrophobic contact with the protein surface.

The Grb2 SH3C domain binds to two different epitopes in Gab2, each with an RxxK core motif within a variable proline-rich sequence. A structural study has elucidated the precise molecular interactions involved in Grb2–Gab2 binding (Figure 1-8).[38] The Gab2a epitope presents the Arg and Lys residues within a polyproline II helix conformation, while in the Gab2b epitope they are found within a 3_{10} helix. Gab2b binds approximately 20-fold stronger than Gab2a (3 μM vs. 58 μM), but crystal structures of both peptides bound to Grb2 SH3C show very similar binding orientations (Figure 1-9). In both cases, essential hydrogen bonds are formed to Glu13 and Glu16 on the Grb2 protein, though due to the different length of turns in 3_{10} and PPII helices (6 Å vs 9 Å), the cationic sidechains are projected at different angles towards the Glu sidechains. Significant hydrophobic contacts stabilize both interactions, so rationalizing the difference in binding affinity for the two peptides is difficult. Mutagenesis experiments indicate that replacing the Arg or Lys residues in either sequence with alanine results in unquantifiably weak binding.

Figure removed for copyright reasons

Figure 1-9: Details of the binding of the Gab2b and Gab2a peptides to Grb2 SH3C. The upper panels show the electrostatic interactions between the Gab2b 3_{10} peptide (A) and Gab2a PPII peptide (B) binding to Grb2 SH3C. In both cases, hydrogen bonds are made to Glu16, but the peptides approach in different orientations due to the difference in length of the two helices. The bottom panels depict the full crystal structures of the interactions. Figure from Harkiolaki *et al*.[38]

1.4. Inhibition of SH3 domain interactions

The inhibition of PPIs has proved challenging for traditional medicinal chemistry efforts. Unlike for enzymes, which have well-defined active sites, PPIs occur along large, relatively featureless surfaces.[81,82] Inhibition has been made tractable by the discovery of "hot spots": individual residues or small groups of residues which contribute the majority of favourable free energy of binding of a PPI.[83] By carefully targeting these key regions of the interaction surface, even large PPIs have been successfully inhibited.[84,85] Particularly for promiscuous SH3 domains, the problem of affinity is complicated by additional selectivity concerns. Nonetheless, diverse approaches have been attempted to search for inhibitors.

1.4.1. Peptide ligands

By selectively mutating residues of the 3BP1 peptide (APTMPPPLPP), the Serrano group induced selectivity for particular SH3 domains. While the peptide binds with similar

affinity to the Abl- and Fyn-SH3 domains, introducing mutations (M4Y, P5S, L8P) based on computer modelling increased affinity for Abl-SH3 20-fold while decreasing affinity for Fyn-SH3 10-fold.[86] The study showed that the first six amino acids of 3BP1 are important for determining selectivity, while the remaining four are important for affinity. The authors rationalized that differences in affinity are due not only to specific sidechain-sidechain interactions, but also to the different entropic costs of adopting a PPII conformation. Selectivity resulted from the interactions between residues four and five of the peptide with the RT and n-Src loops. It was also found that in principle, every Pro residue could be substituted without causing a large loss of binding, provided that a suitable replacement was found for each site.

1.4.2. Peptide dimers

Inhibition may be dramatically enhanced *via* the simultaneous binding of two domains with a single molecule, an approach exemplified by Cussac *et al.* in binding both SH3 domains of Grb2.[87] A dimeric peptide made of two identical proline-rich sequences from Sos linked by a lysine spacer showed an increase in binding affinity from 16 μM to 40 nM when compared to the monomeric peptide (Figure 1-10). The dimer was selective for Grb2 over PI3K and Nck, and could be modified to enter cells where it showed antiproliferative activity. Further studies showed that the compound inhibits the growth of HER2 cancer cells and has antitumor effects in human cancer xenografts.[80] The group went on to show that the compound is a potential therapeutic agent for chronic myelogenous leukaemia (CML).[88] A theoretical model of multivalent binding to SH3 domains predicts this dimeric effect to be general, with significant binding enhancement possible if each ligand has at least mM affinity.[89]

Figure removed for copyright reasons

Figure 1-10: Left: two identical proline-rich peptides linked with a lysine spacer to give a dimer capable of simultaneously binding to two domains. Right: Model of the Grb2-peptide dimer complex showing the PPII helices (blue) and lysine linker (yellow) complexed with the *N*- and *C*-SH3 domains (red) of Grb2. The SH2 domain is in green. Copyright © 1998 by the Federation of American Societies for Experimental Biology.[87]

1.4.3. Combinatorial approaches

The Schreiber group pioneered split-pool combinatorial approaches in the search for SH3 binders. Initial efforts screened a biased library of two million peptides of the form xxxPPxPxx, which contains a known binding sequence.[39] Many residues were found to be highly conserved, which led to the synthesis of the peptide RKLPPRPRR which binds the PI3K SH3 domain with a K_d of 7.6 μM—at that time comparable to the strongest known binders. Further screening in which non-peptidic elements were conjugated to peptides gave a molecule (Figure 1-11a) that bound Src SH3 with a K_d of 3.4 μM. When measured against the PI3K SH3, this molecule displayed a 52-fold increase in selectivity compared with the best known peptide.[90] Multidimensional NMR spectroscopy of these complexes showed that the non-peptide portions of the ligands interact with the specificity pocket of Src SH3 differently from peptides,[91] although not in all cases.[92] Generally peptide elements form contacts with the n-Src loop whereas non-peptide elements contact the RT loop of the same pocket. In both cases there are interactions with the tyrosine-based floor.[93] Having succeeded in attaining high binding affinities, the group focused on achieving selectivity

between highly conserved domains.[94] Following the same method of split-pool synthesis it was possible to select an analogue of Figure 1-11 that displayed 165-fold greater affinity for Hck over PI3K.

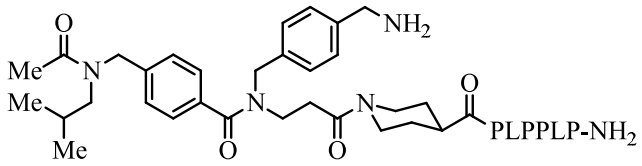


Figure 1-11: Structure of a Scr inhibitor in which non-peptide elements are conjugated with a peptide. This molecule has an affinity of 3.4 μM for Src and shows good selectivity over PI3K.

1.4.4. N-Substituted peptoids

Nguyen *et al*. were the first to explain the SH3 domain's selectivity for proline residues as a preference for *N*-substituted residues, of which proline is the most common endogenous example.[44,95] Replacing the key prolines of a peptide with *N*-substituted residues gave a Grb2 SH3N binder with 100-fold greater affinity. The authors screened a series of SH3 ligands in which each of the two required PxxP prolines was replaced by a series of 11 non-natural *N*-substituted glycines against Sem5, Crk, Grb2 and Src; more than half of these peptoids bound as well as, or better than, the natural peptides. The group went on to show that these peptoids could be selectively tuned to target a single SH3 domain in a test set, with orthogonal selectivity for Crk SH3 against Grb2 and Src.[96] The groups of Tuchscherer and Feller developed this approach (Figure 1-12) to give pseudoproline-containing peptides able to inhibit Grb2 SH3/SoS complex formation in a competition assay using whole-cell lysates.[97] In a unified strategy, Vidal *et al*. incorporated the *N*-alkyl glycine amino acids used by Nguyen into their previously reported proline-rich peptide dimers (Section 1.4.2).[98] The resultant lysine-linked peptoid, VPPPV(peG)PRRR)$_2$K (peG: (*S*)-(α-phenyl) ethylglycine), gave a K_i of 0.20 nM for the Grb2 SH3 domains, compared to 38 nM when peG is replaced with proline. While a crystal structure of the complex was not obtained,

molecular modelling studies suggest that enhanced binding affinity is due to aromatic contacts between the phenyl group of the non-natural amino acid and the binding surface.

Figure 1-12: Representation of a PPII helix binding to an SH3 domain. Proline is required for the ligand to adopt the necessary PPII structure, but due to a lack of functional groups on proline those residues do not have optimal shape complementarity for binding. Incorporation of a functionalized psuedoproline (modified groups shown in red) allows for better packing and thus improvement of binding affinity. Figure from Tuchscherer, G. *et al*.,[97] copyright © 2001 by WILEY-VCH Verlag GmbH, Weinheim, Fed. Rep. of Germany.

1.4.5. Small-molecule inhibitors

A combination of virtual and experimental screening allowed Betzi *et al*. to identify drug-like molecules that functionally target the HIV-1 Nef SH3 binding surface.[99] The most promising compound, DLC27 (Figure 1-15a), displayed an apparent dissociation constant of 980 nM in ITC experiments, and was shown by NMR HSQC experiments to interact with Nef at the same position as the endogenous ligand.

With the mouse Tec Kinase SH3 domain as a model system for structure-based ligand design, Inglis *et al*. used computational methods to identify several simple heterocyclic inhibitors, including a series of quinolines.[100,101] Using NMR chemical shift perturbations, structure-activity relationships, and site-directed mutagenesis, the best compound, a 6-substituted-2-aminoquinoline (Figure 1-15b), was shown to bind competitively with a proline-rich peptide with a K_d of 22 μM. Preliminary studies showed mixed selectivity, with some of these quinolines unable to bind to Fyn, binding Hck weakly, and with approximately equal affinities for Tec and Nck. Further SAR aided the development of a

related compound (Figure 1-15d) with a K_d of 9 μM, where the physiologically incompatible acetal functionality is replaced with a heterocycle.[102] The group went on to synthesize a series of *N*-benzyl derivatives of these compounds,[103] and while binding affinity was reduced it furthered a model for the specific interactions of the quinoline-Tec complex (Figure 1-13).

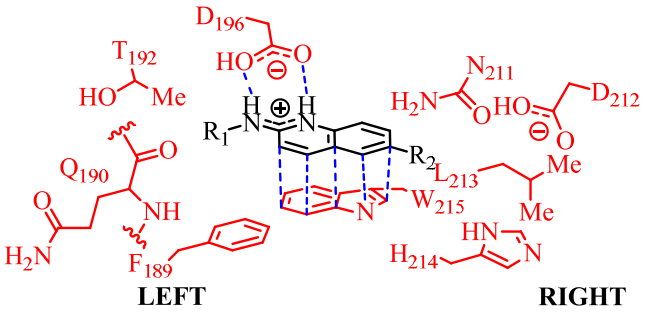


Figure 1-13: Model for mechanism of binding of 2-aminoquinoline to the Tec SH3 domain, in which residues to the left and right of the binding site make contacts with the aminoquinoline substituents. The nature of R^1 is especially important as it impacts upon the salt bridge formed with D196.

Structurally related benzoquinoline molecules were discovered by computational docking and validated by a fluorescence polarization-based assay as micromolar inhibitors for the Src-SH3 domain (Figure 1-15e).[104] Based on this *in silico* work, a binding model for Figure 1-15e was proposed in which both hydrophobic and hydrogen bonding contacts are critical (Figure 1-14).

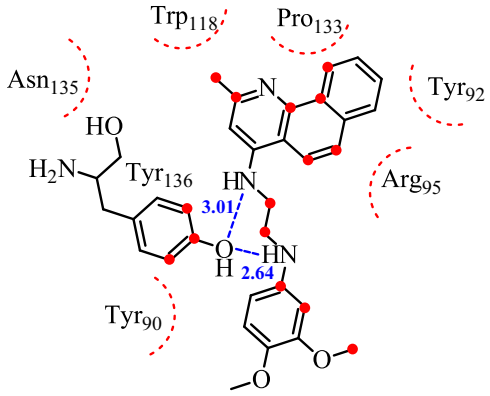


Figure 1-14: A representation of the hydrophobic (red) and hydrogen bonding contacts (blue) in the complex of a benzoquinoline (see Figure 1-15e) with Src SH3, calculated using LIGPLOT (distances are in Å).[105]

Using a yeast-based high-throughput assay, the Sharma group identified UCS15A, produced by *Streptomyces,* as an SH3 domain inhibitor.[106] Inhibition was observed for the interaction of Sam68, a nuclear RNA-binding protein, with a wide variety of SH3 domains including those of Src, Grb2 and PLCγ, but no effect was seen on SH2-mediated interactions. Analyses suggested the compound did not bind to the SH3 domain, but rather interacted directly with the proline-rich regions on target proteins.[107] The nature of protein binding, was not elucidated, though large discrepancies between *in vitro* and *in vivo* studies were observed in that much higher doses were required in the former case. Inhibition was not seen when the SH3 bound complex was dosed, but only when Sam68 was preincubated with UCS15A, hinting at a possible conformational change in the proline-rich region. The Sharma group went on to synthesize greatly simplified structures that showed more potent activity and proved to be less cytotoxic (Figure 1-15f).[108]

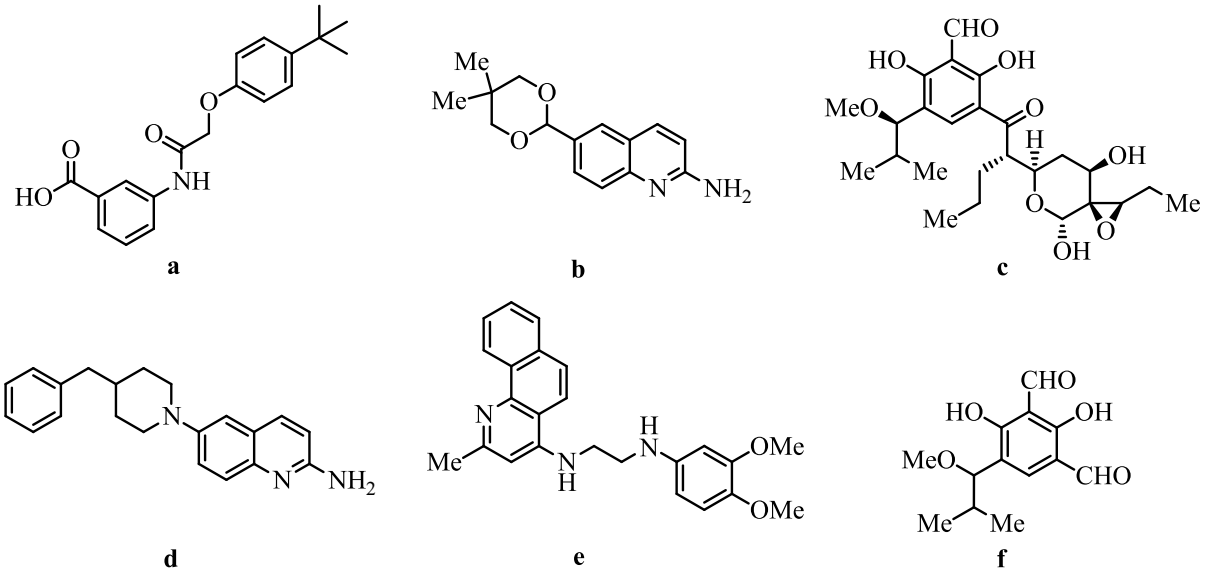


Figure 1-15: Small-molecule SH3 inhibitors. (a) DLC27, an HIV-1 Nef binder, (b) A 6-substituted-2-aminoquinoline targeting the Tec kinase, (c) Natural product UCS15A, thought to bind proline-rich peptides and resulting in the inhibition of many SH3 families, (d) Derivative of compound (b) with an improved physiological profile, (e) A benzoquinoline Src binder, (f) Structurally simplified synthetic analogue of UCS15A.

1.5. Peptidomimetics for inhibiting PPIs

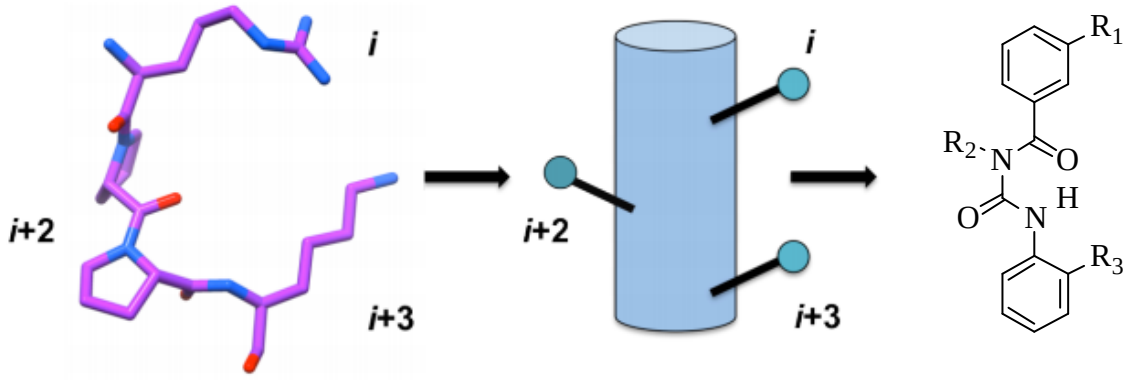


Figure 1-16 Theoretical progression from a peptide to a peptidomimetic. The PPII helix (left) can be construed as a backbone for projecting sidechains (centre). These can be displayed in same manner by a wholly-synthetic scaffold (right).

One recent approach to PPI inhibitors seeks to design molecules which mimic the proteins involved in the interaction. These peptidomimetics display sidechains with the same spatial and angular orientation as proteins, but using non-peptidic elements (Figure 1-16).[81,109,110] Approaches included covalently bonded peptides,[111–113] β-peptides,[114–116] and synthetic small molecules[117–120] have been applied to mimic a variety of protein secondary structures. Among small molecule peptidomimetics, the greatest effort has been devoted to mimics of α-helices. The majority of the free energy of binding of helical protein domains comes from sidechain, as opposed to backbone, interactions,[121,122] thus a wide variety of scaffolds have proved useful (Figure 1-17). There is no clear consensus on an "ideal" peptidomimetic since all such molecules necessarily simplify the actual protein structure. Concerns such as ease of synthesis, adaptability to different sidechains, and the number of sidechains to be mimicked all affect the choice of a peptidomimetic scaffold.

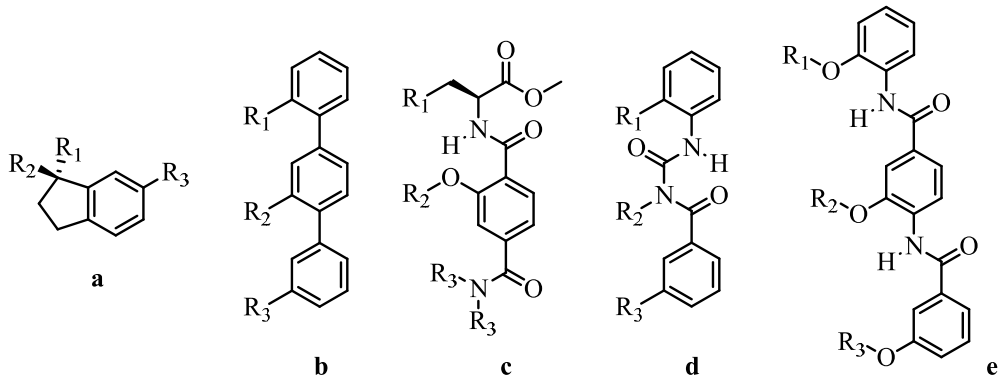


Figure 1-17 Peptidomimetic scaffolds. (a) indane, (b) terphenyl, (c) terephthalamide, (d) benzoylurea, (e) benzamide.[123]

One inescapable objective, however, is that the conformation of the peptidomimetic matches that of the protein. Certain scaffolds have been designed to incorporate a great deal of flexibility, allowing a single molecule to mimic multiple secondary structures *via* an induced fit mechanism.[119,124] Flexibility, however, adds an entropic penalty to binding, which may cancel otherwise favourable interactions. Minimizing this entropic penalty requires a relatively constrained molecule which targets a particular protein structure. While this approach is less general and requires more care in the design phase, it holds promise for discovering more potent and selective inhibitors of PPIs.

1.5.1. Intramolecular hydrogen bonding in peptidomimetics

While covalently-constrained systems allow for precisely-tuned structures, they may be synthetically inaccessible or suffer from poor solubility. As an alternative strategy, a peptidomimetic can instead be constrained by an intramolecular interaction such as a hydrogen bond. Although the nature of hydrogen bonding has been debated in the literature since at least the 1930s,[125] IUPAC issued a technical definition of the interaction in 2011, prefacing it with:

> The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X—H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation.[126]

The definition then proceeds to list various forms of structural, spectroscopic, and computational evidence that may indicate whether a hydrogen bond is present. Additionally, large-scale database searches have elucidated the favoured conformations and energies of hydrogen bonds in published crystal structures.[127,128] In general, hydrogen bond strength increases as the angle between the donor and acceptor tends towards 180º, and as the acidity of the H increases. Intramolecular hydrogen bonds are most favourable when six-membered rings are formed,[127] although larger and smaller rings can still be favourable. Numerous peptidomimetic scaffolds, including the terephthalamide, benzoylurea, and benzamide (Figure 1-17c-e) form such intramolecular rings to direct conformation, allowing the molecules to remain more polar than covalently-bonded structures such as the terphenyl (Figure 1-17b). By modifying the hydrogen bonding pattern of the system, mimics have been prepared for both α-helices[117,118,120] and β-strands,[129] and could potentially be adapted for other systems.

1.6. Conclusions

Nearly 25 years of studies have uncovered the pivotal role of SH3 domains in a wide range of biological processes. As only a few key residues appear sufficient to define the fold, diverse changes are possible within the binding domain, allowing SH3 to play a versatile role in PPIs. Proline-rich peptide ligands bind in a well-understood manner and show moderate affinities, but display poor selectivity within SH3 families. Adding non-peptidic appendages allows the development of compounds with greater affinities and selectivities since a greater number of binding sites are utilized. Despite the efforts of many researchers, the number of classes of small-molecule inhibitors remains disappointing, hinting that new approaches are required. This thesis aims to extend the field of peptidomimetics to incorporate the types of structures required to inhibit SH3 domain

interactions. Using Grb2 SH3C as a target, computational, synthetic, and spectroscopic methods will be applied to discover novel inhibitors.

1.7. References

(1) Copley, R. R. The animal in the genome: comparative genomics and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2008**, *363*, 1453–1461.

(2) Pennisi, E. Working the (Gene Count) Numbers: Finally, a Firm Answer? *Science* **2007**, *316*, 1113–1113.

(3) Hillier, L. W.; Coulson, A.; Murray, J. I.; Bao, Z.; Sulston, J. E.; Waterston, R. H. Genomics in C. elegans: So many genes, such a little worm. *Genome Res.* **2005**, *15*, 1651–1660.

(4) Stumpf, M. P. H.; Thorne, T.; de Silva, E.; Stewart, R.; An, H. J.; Lappe, M.; Wiuf, C. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6959–6964.

(5) Simonis, N.; Rual, J.-F.; Carvunis, A.-R.; Tasan, M.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Sahalie, J. M.; Venkatesan, K.; Gebreab, F.; Cevik, S.; Klitgord, N.; Fan, C.; Braun, P.; Li, N.; Ayivi-Guedehoussou, N.; Dann, E.; Bertin, N.; Szeto, D.; Dricot, A.; Yildirim, M. A.; Lin, C.; de Smet, A.-S.; Kao, H.-L.; Simon, C.; Smolyar, A.; Ahn, J. S.; Tewari, M.; Boxem, M.; Milstein, S.; Yu, H.; Dreze, M.; Vandenhaute, J.; Gunsalus, K. C.; Cusick, M. E.; Hill, D. E.; Tavernier, J.; Roth, F. P.; Vidal, M. Empirically-controlled mapping of the Caenorhabditis elegans protein-protein interactome network. *Nat. Methods* **2009**, *6*, 47–54.

(6) Pawson, T.; Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **2000**, *14*, 1027–1047.

(7) David L. Nielson; Michael M. Cox *Lehninger Principles of Biochemistry*; 3rd, Third Edition, First Printing.; Worth, 2000.

(8) Wetlaufer, D. B. Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 697–701.

(9) Campbell, I. D.; Downing, A. K. Building protein structure and function from modular units. *Trends Biotechnol.* **1994**, *12*, 168–172.

(10) Flynn, D. C. Adaptor proteins. *Oncogene* **2001**, *20*, 6270–6272.

(11) Apic, G.; Gough, J.; Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **2001**, *310*, 311–325.

(12) Dalgarno, D. C.; Botfield, M. C.; Rickles, R. J. SH3 domains and drug design: Ligands, structure, and biological function. *Peptide Sci.* **1997**, *43*, 383–400.

(13) Sawyer, T.; Boyce, B.; Dalgarno, D.; Iuliucci, J. Src inhibitors: genomics to therapeutics. *Expert Opin. Investig. Drugs* **2001**, *10*, 1327–1344.

(14) Vidal, M.; Gigoux, V.; Garbay, C. SH2 and SH3 domains as targets for anti-proliferative agents. *Crit. Rev. Oncol. Hematol.* **2001**, *40*, 175–186.

(15) Feller, S. M.; Tuchscherer, G.; Voss, J. High affinity molecules disrupting GRB2 protein complexes as a therapeutic strategy for chronic myelogenous leukaemia. *Leuk. Lymphoma* **2003**, *44*, 411–427.

(16) Lu, X.-L.; Cao, X.; Liu, X.-Y.; Jiao, B.-H. Recent Progress of Src SH2 and SH3 Inhibitors as Anticancer Agents. *Curr. Med. Chem.* **2010**, *17*, 1117–1124.

(17) Freund, C.; Schmalz, H.-G.; Sticht, J.; Kühne, R. Proline-Rich Sequence Recognition Domains (PRD): Ligands, Function and Inhibition. In *Protein-Protein*

Interactions as New Drug Targets; Klussmann, E.; Scott, J., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; Vol. 186, pp. 407–429.

(18) Luccarelli, J.; Thompson, S.; Hamilton, A. D. SH3 domains as drug targets. In *Protein-Protein Interactions in Drug Discovery*; Dömling, A., Ed.; Methods and Principles in Medicinal Chemistry; Wiley-VCH, 2013; pp. 101–128.

(19) Mayer, B. J.; Hamaguchi, M.; Hanafusa, H. A novel viral oncogene with structural similarity to phospholipase C. *Nature* **1988**, *332*, 272–275.

(20) Stahl, M. L.; Ferenz, C. R.; Kelleher, K. L.; Kriz, R. W.; Knopf, J. L. Sequence similarity of phospholipase C with the non-catalytic region of src. *Nature* **1988**, *332*, 269–272.

(21) Kärkkäinen, S.; Hiipakka, M.; Wang, J.-H.; Kleino, I.; Vähä-Jaakkola, M.; Renkema, G. H.; Liss, M.; Wagner, R.; Saksela, K. Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. *EMBO Rep.* **2006**, *7*, 186–191.

(22) Finn, R. D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E. L. L.; Eddy, S. R.; Bateman, A. The Pfam protein families database. *Nucleic Acids Res.* **2009**, *38*, D211–D222.

(23) Bilwes, A. M.; Alex, L. A.; Crane, B. R.; Simon, M. I. Structure of CheA, a Signal-Transducing Histidine Kinase. *Cell* **1999**, *96*, 131–141.

(24) Edmondson, S. P.; Turri, J.; Smith, K.; Clark, A.; Shriver, J. W. Structure, Stability, and Flexibility of Ribosomal Protein L14e from Sulfolobus Solfataricus. *Biochemistry* **2009**, *48*, 5553–5562.

(25) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **1992**, *359*, 851–855.

(26) Yu, H.; Rosen, M.; Shin, T.; Seidel-Dugan, C.; Brugge, J.; Schreiber, S. Solution structure of the SH3 domain of Src and identification of its ligand-binding site. *Science* **1992**, *258*, 1665 –1668.

(27) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(28) Larson, S. M.; Davidson, A. R. The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. *Protein Sci.* **2000**, *9*, 2170–2180.

(29) Di Nardo, A. A.; Larson, S. M.; Davidson, A. R. The Relationship Between Conservation, Thermodynamic Stability, and Function in the SH3 Domain Hydrophobic Core. *J. Mol. Biol.* **2003**, *333*, 641–655.

(30) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.

(31) Ren, R.; Mayer, B.; Cicchetti, P.; Baltimore, D. Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **1993**, *259*, 1157 –1161.

(32) Kay, B. K.; Williamson, M. P.; Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *Faseb J.* **2000**, *14*, 231 –241.

(33) MacArthur, M. W.; Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **1991**, *218*, 397–412.

(34) Ball, L. J.; Kühne, R.; Schneider-Mergener, J.; Oschkinat, H. Recognition of Proline-Rich Motifs by Protein-Protein-Interaction Domains. *Angew. Chem. Int. Ed.* **2005**, *44*, 2852–2869.
(35) Adzhubei, A. A.; Sternberg, M. J. E. Left-handed Polyproline II Helices Commonly Occur in Globular Proteins. *J. Mol. Biol.* **1993**, *229*, 472–493.
(36) Creamer, T. P.; Campbell, M. N. Determinants of the polyproline II helix from modeling studies. In *Unfolded Proteins*; Academic Press, 2002; Vol. 62, pp. 263–282.
(37) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A.-F.; Creamer, T. P. Host−Guest Study of Left-Handed Polyproline II Helix Formation. *Biochemistry* **2001**, *40*, 14376–14383.
(38) Harkiolaki, M.; Tsirka, T.; Lewitzky, M.; Simister, P. C.; Joshi, D.; Bird, L. E.; Jones, E. Y.; O'Reilly, N.; Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the SH3C Domain of Grb2. *Structure* **2009**, *17*, 809–822.
(39) Chen, J. K.; Lane, W. S.; Brauer, A. W.; Tanaka, A.; Schreiber, S. L. Biased combinatorial libraries: novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. *J. Am. Chem. Soc.* **1993**, *115*, 12591–12592.
(40) Rozakis-Adcock, M.; Fernley, R.; Wade, J.; Pawson, T.; Bowtell, D. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. *Nature* **1993**, *363*, 83–85.
(41) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **1992**, *359*, 851–855.
(42) Lim, W. A.; Richards, F. M.; Fox, R. O. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **1994**, *372*, 375–379.
(43) Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. L. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **1994**, *266*, 1241–1247.
(44) Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A. Exploiting the Basis of Proline Recognition by SH3 and WW Domains: Design of N-Substituted Inhibitors. *Science* **1998**, *282*, 2088 –2092.
(45) Qiu, M.; Hua, S.; Agrawal, M.; Li, G.; Cai, J.; Chan, E.; Zhou, H.; Luo, Y.; Liu, M. Molecular Cloning and Expression of HumanGrap-2, a Novel Leukocyte-Specific SH2- and SH3-Containing Adaptor-like Protein That Binds toGab-1. *Biochem. Biophys. Res. Com.* **1998**, *253*, 443–447.
(46) Berry, D. M.; Nash, P.; Liu, S. K.-W.; Pawson, T.; McGlade, C. J. A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling. *Curr. Biol.* **2002**, *12*, 1336–1341.
(47) Liu, Q.; Berry, D.; Nash, P.; Pawson, T.; McGlade, C. J.; Li, S. S.-C. Structural basis for specific binding of the Gads SH3 domain to an RxxK motif-containing SLP-76 peptide: a novel mode of peptide recognition. *Mol. Cell* **2003**, *11*, 471–481.
(48) Harkiolaki, M.; Lewitzky, M.; Gilbert, R. J. C.; Jones, Ey.; Bourette, R. P.; Mouchiroud, G.; Sondermann, H.; Moarefi, I.; Feller, S. M. Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76. *EMBO J.* **2003**, *22*, 2571–2582.
(49) Mayer, B. SH3 domains: complexity in moderation. *J. Cell Sci.* **2001**, *114*, 1253–1263.
(50) Ladbury, J. E.; Arold, S. Searching for specificity in SH domains. *Chem. Biol.* **2000**, *7*, R3–R8.
(51) Posas, F.; Saito, H. Osmotic Activation of the HOG MAPK Pathway via Ste11p MAPKKK: Scaffold Role of Pbs2p MAPKK. *Science* **1997**, *276*, 1702 –1705.

(52) Zarrinpar, A.; Park, S.-H.; Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **2003**, *426*, 676–680.
(53) Seet, B. T.; Berry, D. M.; Maltzman, J. S.; Shabason, J.; Raina, M.; Koretzky, G. A.; McGlade, C. J.; Pawson, T. Efficient T-cell receptor signaling requires a high-affinity interaction between the Gads C-SH3 domain and the SLP-76 RxxK motif. *EMBO J.* **2007**, *26*, 678–689.
(54) Koch, C. A.; Anderson, D.; Moran, M. F.; Ellis, C.; Pawson, T. SH2 and SH3 Domains: Elements That Control Interactions of Cytoplasmic Signaling Proteins. *Science* **1991**, *252*, 668–674.
(55) Lowenstein, E. J.; Daly, R. J.; Batzer, A. G.; Li, W.; Margolis, B.; Lammers, R.; Ullrich, A.; Skolnik, E. Y.; Bar-Sagi, D.; Schlessinger, J. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **1992**, *70*, 431–442.
(56) Cheng, A. M.; Saxton, T. M.; Sakai, R.; Kulkarni, S.; Mbamalu, G.; Vogel, W.; Tortorice, C. G.; Cardiff, R. D.; Cross, J. C.; Muller, W. J.; Pawson, T. Mammalian Grb2 Regulates Multiple Steps in Embryonic Development and Malignant Transformation. *Cell* **1998**, *95*, 793–803.
(57) Dharmawardana, P. G.; Peruzzi, B.; Giubellino, A.; Burke, T. R., Jr; Bottaro, D. P. Molecular targeting of growth factor receptor-bound 2 (Grb2) as an anti-cancer strategy. *Anticancer Drugs* **2006**, *17*, 13–20.
(58) Feller, S. M.; Lewitzky, M. Potential disease targets for drugs that disrupt protein--protein interactions of Grb2 and Crk family adaptors. *Curr. Pharm. Des.* **2006**, *12*, 529–548.
(59) Giubellino, A.; Burke, T. R.; Bottaro, D. P. Grb2 Signaling in Cell Motility and Cancer. *Expert Opin. Ther. Targets* **2008**, *12*, 1021–1033.
(60) Huebner, K.; Kastury, K.; Druck, T.; Salcini, A. E.; Lanfrancone, L.; Pelicci, G.; Lowenstein, E.; Li, W.; Park, S. H.; Cannizzaro, L. Chromosome locations of genes encoding human signal transduction adapter proteins, Nck (NCK), Shc (SHC1), and Grb2 (GRB2). *Genomics* **1994**, *22*, 281–287.
(61) Daly, R. J.; Binder, M. D.; Sutherland, R. L. Overexpression of the Grb2 gene in human breast cancer cell lines. *Oncogene* **1994**, *9*, 2723–2727.
(62) Verbeek, B. S.; Adriaansen-Slot, S. S.; Rijksen, G.; Vroom, T. M. Grb2 overexpression in nuclei and cytoplasm of human breast cells: a histochemical and biochemical study of normal and neoplastic mammary tissue specimens. *J. Pathol.* **1997**, *183*, 195–203.
(63) Dankort, D.; Maslikowski, B.; Warner, N.; Kanno, N.; Kim, H.; Wang, Z.; Moran, M. F.; Oshima, R. G.; Cardiff, R. D.; Muller, W. J. Grb2 and Shc Adapter Proteins Play Distinct Roles in Neu (ErbB-2)-Induced Mammary Tumorigenesis: Implications for Human Breast Cancer. *Mol. Cell. Biol.* **2001**, *21*, 1540–1551.
(64) Yang, S. S.; Van Aelst, L.; Bar-Sagi, D. Differential interactions of human Sos1 and Sos2 with Grb2. *J. Biol. Chem.* **1995**, *270*, 18212–18215.
(65) Lewitzky, M.; Kardinal, C.; Gehring, N. H.; Schmidt, E. K.; Konkol, B.; Eulitz, M.; Birchmeier, W.; Schaeper, U.; Feller, S. M. The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif. *Oncogene* **2001**, *20*, 1052–1062.
(66) Simister, P. C.; Feller, S. M. Order and disorder in large multi-site docking proteins of the Gab family—implications for signalling complex formation and inhibitor design strategies. *Mol. BioSyst.* **2012**, *8*, 33–46.
(67) Ainsworth, R.; Bartlett, J. M. S.; Going, J. J.; Mallon, E. A.; Forsyth, A.; Richmond, J.; Angerson, W.; Watters, A.; Dunne, B. IHC for Her2 with CBE356 antibody is a

more accurate predictor of Her2 gene amplification by FISH than HercepTest(TM) in breast carcinoma. *J Clin Pathol* **2005**, *58*, 1086–1090.

(68) Slamon, D. J.; Clark, G. M.; Wong, S. G.; Levin, W. J.; Ullrich, A.; McGuire, W. L. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **1987**, *235*, 177–182.

(69) Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; Baselga, J.; Norton, L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **2001**, *344*, 783–792.

(70) Baselga, J. Adjuvant Trastuzumab: A Milestone in the Treatment of HER-2-Positive Early Breast Cancer. *Oncologist* **2006**, *11*, 4–12.

(71) Nahta, R.; Yu, D.; Hung, M.-C.; Hortobagyi, G. N.; Esteva, F. J. Mechanisms of Disease: understanding resistance to HER2-targeted therapy in human breast cancer. *Nat. Clin. Pract. Oncol.* **2006**, *3*, 269–280.

(72) Geyer, C. E.; Forster, J.; Lindquist, D.; Chan, S.; Romieu, C. G.; Pienkowski, T.; Jagiello-Gruszfeld, A.; Crown, J.; Chan, A.; Kaufman, B.; Skarlos, D.; Campone, M.; Davidson, N.; Berger, M.; Oliva, C.; Rubin, S. D.; Stein, S.; Cameron, D. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N. Engl. J. Med.* **2006**, *355*, 2733–2743.

(73) Baselga, J.; Cortés, J.; Kim, S.-B.; Im, S.-A.; Hegg, R.; Im, Y.-H.; Roman, L.; Pedrini, J. L.; Pienkowski, T.; Knott, A.; Clark, E.; Benyunes, M. C.; Ross, G.; Swain, S. M. Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer. *N. Engl. J. Med.* **2012**, *366*, 109–119.

(74) Phillips, G. D. L.; Li, G.; Dugger, D. L.; Crocker, L. M.; Parsons, K. L.; Mai, E.; Blättler, W. A.; Lambert, J. M.; Chari, R. V. J.; Lutz, R. J.; Wong, W. L. T.; Jacobson, F. S.; Koeppen, H.; Schwall, R. H.; Kenkare-Mitra, S. R.; Spencer, S. D.; Sliwkowski, M. X. Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody–Cytotoxic Drug Conjugate. *Cancer Res.* **2008**, *68*, 9280–9290.

(75) Girish, S.; Gupta, M.; Wang, B.; Lu, D.; Krop, I. E.; Vogel, C. L.; Iii, H. A. B.; LoRusso, P. M.; Yi, J.-H.; Saad, O.; Tong, B.; Chu, Y.-W.; Holden, S.; Joshi, A. Clinical pharmacology of trastuzumab emtansine (T-DM1): an antibody–drug conjugate in development for the treatment of HER2-positive cancer. *Cancer Chemother. Pharmacol.* **2012**, *69*, 1229–1240.

(76) Verma, S.; Miles, D.; Gianni, L.; Krop, I. E.; Welslau, M.; Baselga, J.; Pegram, M.; Oh, D.-Y.; Diéras, V.; Guardino, E.; Fang, L.; Lu, M. W.; Olsen, S.; Blackwell, K. Trastuzumab Emtansine for HER2-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **2012**, *367*, 1783–1791.

(77) Stern, H. M. Improving Treatment of HER2-Positive Cancers: Opportunities and Challenges. *Sci. Transl. Med.* **2012**, *4*, 127rv2–127rv2.

(78) Tari, A. M.; Hung, M. C.; Li, K.; Lopez-Berestein, G. Growth inhibition of breast cancer cells by Grb2 downregulation is correlated with inactivation of mitogen-activated protein kinase in EGFR, but not in ErbB2, cells. *Oncogene* **1999**, *18*, 1325–1332.

(79) Bentires-Alj, M.; Gil, S. G.; Chan, R.; Wang, Z. C.; Wang, Y.; Imanaka, N.; Harris, L. N.; Richardson, A.; Neel, B. G.; Gu, H. A role for the scaffolding adapter GAB2 in breast cancer. *Nat. Med.* **2005**, *12*, 114–121.

(80) Gril, B.; Vidal, M.; Assayag, F.; Poupon, M.; Liu, W.; Garbay, C. Grb2-SH3 ligand inhibits the growth of HER2+ cancer cells and has antitumor effects in human cancer xenografts alone and in combination with docetaxel. *Int. J. Cancer* **2007**, *121*, 407–415.

(81) Yin, H.; Hamilton, A. D. Strategies for Targeting Protein-Protein Interactions With Synthetic Agents. *Angew. Chem. Int. Ed.* **2005**, *44*, 4130–4163.
(82) Ross, N. T.; Katt, W. P.; Hamilton, A. D. Synthetic mimetics of protein secondary structure domains. *Proc. R. Soc. A* **2010**, *368*, 989–1008.
(83) Cunningham, B.; Wells, J. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **1989**, *244*, 1081–1085.
(84) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001–1009.
(85) Villoutreix, B. O.; Labbé, C. M.; Lagorce, D.; Laconde, G.; Sperandio, O. A leap into the chemical space of protein-protein interaction inhibitors. *Curr. Pharm. Des.* **2012**, *18*, 4648–4667.
(86) Pisabarro, M. T.; Serrano, L. Rational Design of Specific High-Affinity Peptide Ligands for the Abl-SH3 Domain. *Biochemistry* **1996**, *35*, 10634–10640.
(87) Cussac, D.; Vidal, M.; Leprince, C.; Liu, W.; Tiraboschi, G.; Roques, B.; Garbay, C. A Sos-derived peptidimer blocks the Ras signaling pathway by binding both Grb2 SH3 domains and displays antiproliferative activity. *FASEB J.* **1999**, *13*, 31–39.
(88) Ye, Y.-B.; Lin, J.-Y.; Chen, Q.; Liu, F.; Chen, H.-J.; Li, J.-Y.; Liu, W.-Q.; Garbay, C.; Vidal, M. The cytotoxicity of a Grb2-SH3 inhibitor in Bcr-Abl positive K562 cells. *Biochem. Pharm.* **2008**, *75*, 2080–2091.
(89) Zhou, H.-X. Quantitative relation between intermolecular and intramolecular binding of pro-rich peptides to SH3 domains. *Biophys. J.* **2006**, *91*, 3170–3181.
(90) Combs, A. P.; Kapoor, T. M.; Feng, S.; Chen, J. K.; Daudé-Snow, L. F.; Schreiber, S. L. Protein Structure-Based Combinatorial Chemistry: Discovery of Non-Peptide Binding Elements to Src SH3 Domain. *J. Am. Chem. Soc.* **1996**, *118*, 287–288.
(91) Feng, S.; Kapoor, T. M.; Shirai, F.; Combs, A. P.; Schreiber, S. L. Molecular basis for the binding of SH3 ligands with non-peptide elements identified by combinatorial synthesis. *Chem. Biol.* **1996**, *3*, 661–670.
(92) Morken, J. P.; Kapoor, T. M.; Feng, S.; Shirai, F.; Schreiber, S. L. Exploring the Leucine-Proline Binding Pocket of the Src SH3 Domain Using Structure-Based, Split-Pool Synthesis and Affinity-Based Selection. *J. Am. Chem. Soc.* **1998**, *120*, 30–36.
(93) Feng, S.; Schreiber, S. L. Enantiomeric Binding Elements Interacting at the Same Site of an SH3 Protein Receptor. *J. Am. Chem. Soc.* **1997**, *119*, 10873–10874.
(94) Kapoor, T. M.; Andreotti, A. H.; Schreiber, S. L. Exploring the Specificity Pockets of Two Homologous SH3 Domains Using Structure-Based, Split-Pool Synthesis and Affinity-Based Selection. *J. Am. Chem. Soc.* **1998**, *120*, 23–29.
(95) Aghazadeh, B.; Rosen, M. K. Ligand recognition by SH3 and WW domains: the role of N-alkylation in PPII helices. *Chem. Biol.* **1999**, *6*, R241–246.
(96) Nguyen, J. T.; Porter, M.; Amoui, M.; Miller, W. T.; Zuckermann, R. N.; Lim, W. A. Improving SH3 domain ligand selectivity using a non-natural scaffold. *Chem. Biol.* **2000**, *7*, 463–473.
(97) Tuchscherer, G.; Grell, D.; Tatsu, Y.; Durieux, P.; Fernandez□Carneado, J.; Hengst, B.; Kardinal, C.; Feller, S. Targeting Molecular Recognition: Exploring the Dual Role of Functional Pseudoprolines in the Design of SH3 Ligands. *Angew. Chem. Int. Ed.* **2001**, *40*, 2844–2848.
(98) Vidal, M.; Liu, W.-Q.; Lenoir, C.; Salzmann, J.; Gresh, N.; Garbay, C. Design of peptoid analogue dimers and measure of their affinity for Grb2 SH3 domains. *Biochemistry* **2004**, *43*, 7336–7344.
(99) Betzi, S.; Restouin, A.; Opi, S.; Arold, S. T.; Parrot, I.; Guerlesquin, F.; Morelli, X.; Collette, Y. Protein–protein interaction inhibition (2P2I) combining high throughput

and virtual screening: Application to the HIV-1 Nef protein. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19256 –19261.

(100) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. Identification and Specificity Studies of Small-Molecule Ligands for SH3 Protein Domains. *J. Med. Chem.* **2004**, *47*, 5405–5417.

(101) Inglis, S.; Jones, R.; Fritz, D.; Stojkoski, C.; Booker, G.; Pyke, S. Synthesis of 5-, 6- and 7-substituted-2-aminoquinolines as SH3 domain ligands. *Org. Biomol. Chem.* **2005**, *3*, 2543–2557.

(102) Smith, J. A.; Jones, R. K.; Booker, G. W.; Pyke, S. M. Sequential and Selective Buchwald−Hartwig Amination Reactions for the Controlled Functionalization of 6-Bromo-2-chloroquinoline: Synthesis of Ligands for the Tec Src Homology 3 Domain. *J. Org. Chem.* **2008**, *73*, 8880–8892.

(103) Inglis, S. R.; Jones, R. K.; Booker, G. W.; Pyke, S. M. Synthesis of N-benzylated-2-aminoquinolines as ligands for the Tec SH3 domain. *Bioorg. Med. Chem. Let.* **2006**, *16*, 387–390.

(104) Atatreh, N.; Stojkoski, C.; Smith, P.; Booker, G. W.; Dive, C.; Frenkel, A. D.; Freeman, S.; Bryce, R. A. In silico screening and biological evaluation of inhibitors of Src-SH3 domain interaction with a proline-rich ligand. *Bioorg. Med. Chem. Let.* **2008**, *18*, 1217–1222.

(105) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.

(106) Sharma, S. V.; Oneyama, C.; Yamashita, Y.; Nakano, H.; Sugawara, K.; Hamada, M.; Kosaka, N.; Tamaoki, T. UCS15A, a non-kinase inhibitor of Src signal transduction. *Oncogene* **2001**, *20*, 2068–2079.

(107) Oneyama, C.; Nakano, H.; Sharma, S. V. UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug. *Oncogene* **2002**, *21*, 2037–2050.

(108) Oneyama, C.; Agatsuma, T.; Kanda, Y.; Nakano, H.; Sharma, S. V.; Nakano, S.; Narazaki, F.; Tatsuta, K. Synthetic Inhibitors of Proline-Rich Ligand-Mediated Protein-Protein Interaction: Potent Analogs of UCS15A. *Chem. Biol.* **2003**, *10*, 443–451.

(109) Ross, N. T.; Katt, W. P.; Hamilton, A. D. Synthetic mimetics of protein secondary structure domains. *Philos. Transact. A Math. Phys. Eng. Sci.* **2010**, *368*, 989–1008.

(110) Adler, M. J.; Jamieson, A. G.; Hamilton, A. D. Hydrogen-bonded synthetic mimics of protein secondary structure as disruptors of protein-protein interactions. *Curr. Top. Microbiol. Immunol.* **2011**, *348*, 1–23.

(111) Schafmeister, C. E.; Po, J.; Verdine, G. L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *J. Am. Chem. Soc.* **2000**, *122*, 5891–5892.

(112) Patgiri, A.; Jochim, A. L.; Arora, P. S. A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in α-Helical Conformation. *Acc. Chem. Res.* **2008**, *41*, 1289–1300.

(113) Liu, J.; Wang, D.; Zheng, Q.; Lu, M.; Arora, P. S. Atomic Structure of a Short α-Helix Stabilized by a Main Chain Hydrogen-Bond Surrogate. *J. Am. Chem. Soc.* **2008**, *130*, 4334–4337.

(114) Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H. β-Peptide Foldamers: Robust Helix Formation in a New Family of β-Amino Acid Oligomers. *J. Am. Chem. Soc.* **1996**, *118*, 13071–13072.

(115) Appella, D. H.; Christianson, L. A.; Klein, D. A.; Powell, D. R.; Huang, X.; Barchi, J. J., Jr; Gellman, S. H. Residue-based control of helix shape in beta-peptide oligomers. *Nature* **1997**, *387*, 381–384.
(116) Kritzer, J. A.; Stephens, O. M.; Guarracino, D. A.; Reznik, S. K.; Schepartz, A. beta-Peptides as inhibitors of protein-protein interactions. *Bioorg. Med. Chem.* **2005**, *13*, 11–16.
(117) Rodriguez, J. M.; Hamilton, A. D. Benzoylurea Oligomers: Synthetic Foldamers That Mimic Extended α Helices. *Angew. Chem. Int. Ed.* **2007**, *46*, 8614–8617.
(118) Adler, M. J.; Hamilton, A. D. Oligophenylenaminones as Scaffolds for α-Helix Mimicry. *J. Org. Chem.* **2011**, *76*, 7040–7047.
(119) Ko, E.; Liu, J.; Perez, L. M.; Lu, G.; Schaefer, A.; Burgess, K. Universal Peptidomimetics. *J. Am. Chem. Soc.* **2011**, *133*, 462–477.
(120) Thompson, S.; Vallinayagam, R.; Adler, M. J.; Scott, R. T. W.; Hamilton, A. D. Double-sided α-helix mimetics. *Tetrahedron* **2012**, *68*, 4501–4505.
(121) Stanfield, R. L.; Wilson, I. A. Protein-peptide interactions. *Curr. Op. Struct. Biol.* **1995**, *5*, 103–113.
(122) Koh, J. T.; Cornish, V. W.; Schultz, P. G. An Experimental Approach to Evaluating the Role of Backbone Interactions in Proteins Using Unnatural Amino Acid Mutagenesis. *Biochemistry* **1997**, *36*, 11314–11322.
(123) Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J. Inhibition of α-helix-mediated protein-protein interactions using designed molecules. *Nat. Chem.* **2013**, *5*, 161–173.
(124) Raghuraman, A.; Ko, E.; Perez, L. M.; Ioerger, T. R.; Burgess, K. Pyrrolinone–Pyrrolidine Oligomers as Universal Peptidomimetics. *J. Am. Chem. Soc.* **2011**, *133*, 12350–12353.
(125) Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: Introduction to Modern Structural Chemistry*; 1st edition.; Cornell University Press, 1939.
(126) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* **2011**, *83*, 1637–1641.
(127) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601–2611.
(128) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.
(129) Sutherell, C. L.; Thompson, S.; Scott, R. T. W.; Hamilton, A. D. Aryl-linked imidazolidin-2-ones as non-peptidic β-strand mimetics. *Chem. Commun.* **2012**, *48*, 9834.

2. Virtual Screen for Grb2 Inhibitors

2.1. Introduction

In a first approach towards inhibitors of Grb2, the ability of commercially-available molecules to disrupt the Grb2/Gab2 interaction was investigated. Previous high-throughput and computational screens have identified several SH3 domain inhibitors (*vide supra* Ch.1.4.5),[1–8] indicating that the existing chemical space includes molecules capable of binding to these proteins. To assess as wide a range of molecules as possible, a virtual screening approach was adopted for its speed, low cost, and broad chemical diversity.[9]

Virtual screening attempts to compute the free-energy of binding of small molecules to a protein target. This is a very difficult task as the physics of such interactions are still not fully understood,[10] but methods such as free energy perturbation calculations have shown success in calculating the relative and absolute binding energies of ligand series.[11–13] Nonetheless, achieving accurate and precise results is a great undertaking, involving extensive consideration of initial structure,[14,15] solvation,[13] flexibility,[16] and sampling.[17] This results in considerable computational expense, with the accurate calculation of a single relative free energy of binding requiring days of processor time.[18] These methods are also generally restricted to relatively small perturbations of molecular structure,[19] making them unsuitable for investigating diverse libraries.

Rapid virtual screens ignore a great deal of the complexity of protein-ligand interactions in exchange for speed. In general, the protein structure is treated as fixed, small molecules are not fully flexible, and solvation is limited to simple dielectric models. Rather than using explicit free energy perturbation calculations, binding energies are calculated by largely empirical scoring functions, which consider only a few parameters.[20] Such calculations are

thus wholly unsuitable for accurately ranking compounds,[9] but empirically have been sufficient in numerous cases to find starting points for further molecular optimization.[21,22]

2.2. Virtual ligand selection

While alanine scanning and spot blot analysis of peptide truncates have probed the binding site of Grb2,[23] the chemical space of small molecule inhibitors has not been explored. To maximize the scope of this investigation, a large database of molecules was desired. Given the crude predictive power of virtual screening calculations, even high-ranking molecules are relatively unlikely to be true hits, so to avoid wasted synthesis the search was limited to commercially available compounds. Even the purchasable chemical space, however, is vastly too large to screen, with the database ZINC listing nearly 20 million commercially available compounds.[24,25] To narrow this further, molecules were filtered to comply with Lipinski's "rule of 5,"[26] limiting the search set to molecules with molecular weight < 500, calculated octanol-water partition coefficient of ≤ 5, hydrogen bond donors ≤ 5, and hydrogen bond acceptors ≤ 10. This resulted in a set of ca. 6.3 million molecules (downloaded from ZINC in September 2010). This set included the molecules in their protonation states expected in the range of pH 5.75-8.25, as calculated by ZINC.

2.3. Virtual screen

The docking program GLIDE (version 5.6) was used to conduct the screen.[27,28] The crystal structure of the complex containing the more-tightly binding Gab2b 3_{10} helix (PDB code: 2VWF),[23] was used as the docking target. The protein was processed using GLIDE's default protein preparation workflow: the bound peptide and all water molecules were removed, and the protein structure refined using exhaustive sampling for H-bonding networks. Following a constrained *Impref* minimization with the OPLS-AA 2005 forcefield,[29] optimizing all atoms to an RMSD of 0.3 Å, the receptor grid was generated as a

cube 7 Å in radius around Trp35 in the center of the binding pocket. As the peptide studies confirmed the critical role of a hydrogen bond to the Glu16 sidechain, a constraint was applied requiring all docked molecules to form at least one hydrogen bond to that residue. The molecules were then flexibly docked onto the target protein, using the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD, USA. Binding energies were evaluated using the GlideSP scoring function. The resulting screen ranked each compound by glide-score, which roughly approximates their free energies of binding.

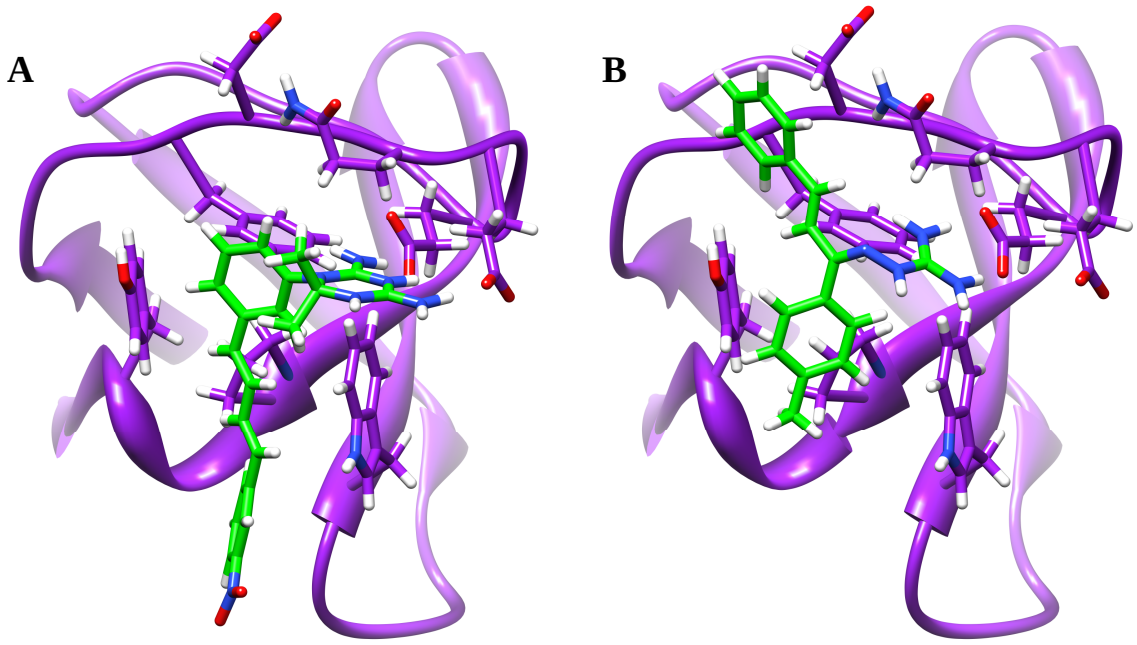


Figure 2-1: Docked poses of the molecules ranked 5 (left) and 164 (right) in the virtual screen bound to the Grb2 SH3C (purple). Both molecules are predicted to make extensive hydrogen-bond and hydrophobic contacts with the protein.

The top 200 virtual hits were screened by eye for binding pose and chemical diversity (Figure 2-1). In general, the molecules had both hydrophobic and hydrophilic components, and consistent with the applied constraint all bound to similar binding pockets on the surface of the protein. Of the top hits, 34 were obtained for *in vitro* testing from the NCI/DTP Open Chemical Repository (Figure 2-2). Samples were obtained as pure solids, which were dissolved as DMSO stocks for further analysis.

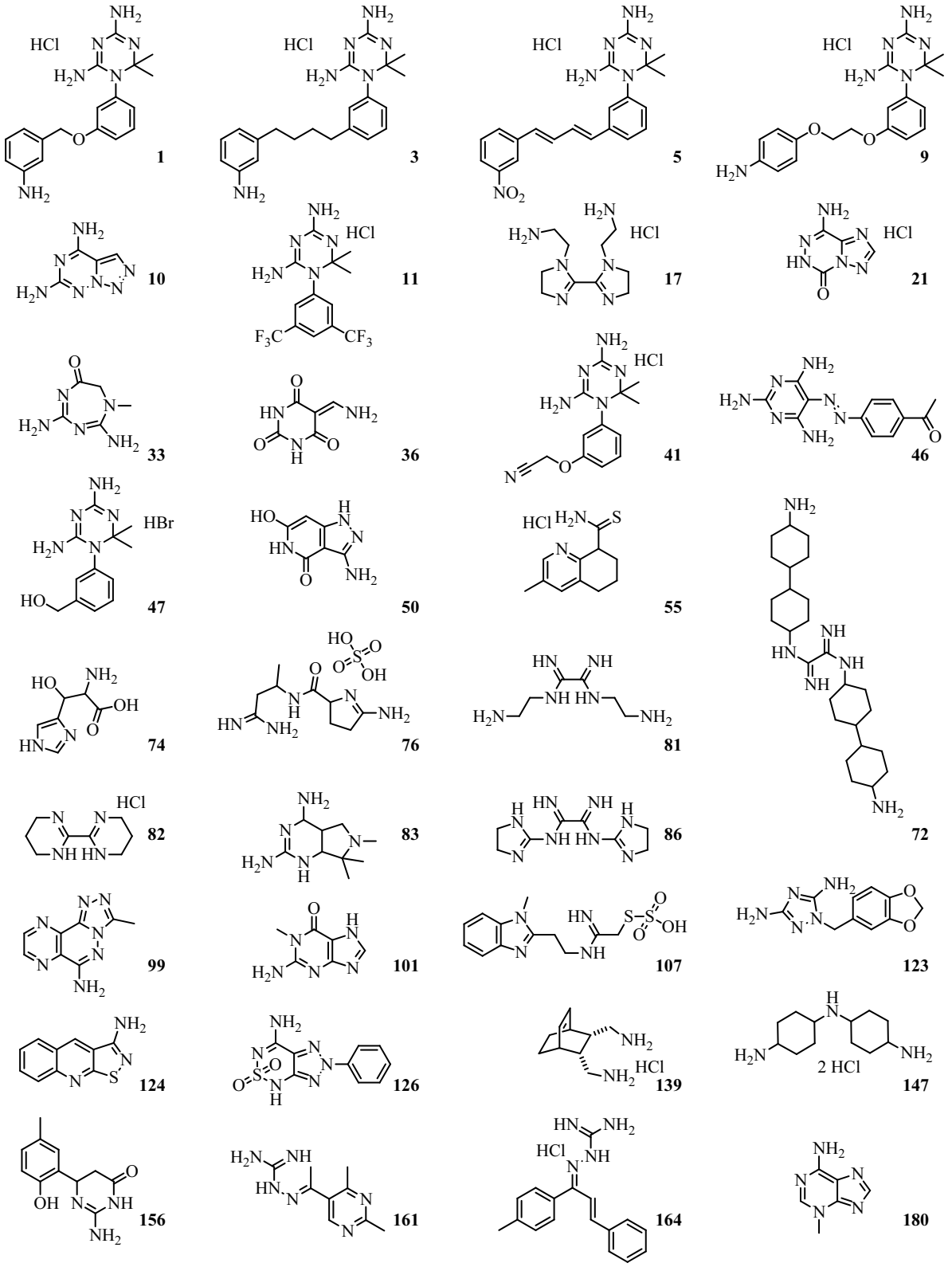


Figure 2-2: Structures of the 34 compounds obtained from the NCI/DTP open chemical repository.

2.4. SPR binding assay

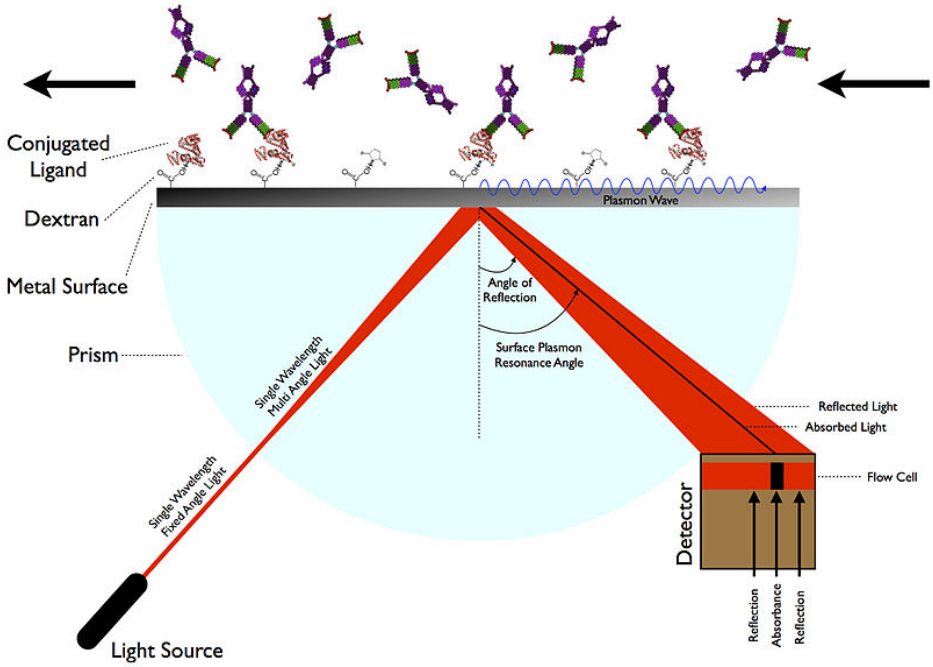


Figure 2-3: Schematic of the SPR experiment. At the SPR wavelength, incident light excites surface plasmons along the metal. Changes to the index of refraction along the interface (in this case, caused by antibody binding), alter the index of refraction, thus changing the SPR angle. This change in angle is detected and plotted to measure the binding kinetics and thermodynamics. Creative commons image from Sari Sabban. [30]

The selected molecules were first tested for binding using surface plasmon resonance (SPR). This technique measures the change of the index of refraction along a metal surface caused by the binding of an analyte.[31,32] In practice, a laser is shined at a thin gold surface adjacent to an aqueous medium. At precise angles, the incoming photons excite an electromagnetic wave along the interface between the metal and the liquid, resulting in a drop in reflected light intensity (Figure 2-3). The angle of maximal excitation (and thus minimal reflectance) is the SPR angle, which varies with the index of refraction of the system. As the metal surface itself is unchanging, the observed change in SPR angle is caused by changes along the metal/liquid interface, in this case caused by the binding of a ligand. The technique, however, measures all changes in refractive index, which could include changes to the buffer or concentration rather than desired binding. Plotting the change in SPR angle with respect to time allows the thermodynamics and kinetics of an

interaction to be calculated.[33] Rather than binding protein directly to the gold surface, a carboxydextran layer is first applied to which the protein is covalently attached. This leaves the protein tethered to the dextran, but otherwise free to rotate and translate.[34] Benchmarking studies have determined that carefully-performed SPR experiments yield reproducible measurements comparable to those obtained by solution experiments (e.g. ITC), indicating that the dextran tether does not generally affect the protein.[35–38]

For this study, wild type Grb2 SH3C was attached to the carboxydextran surface of a CM5 chip *via* amine groups on the protein. The matrix was first activated with a mixture of EDC and NHS, which results in an activated NHS ester. The protein was then dissolved in an acidic (pH 4 acetate) buffer to create an overall positively-charged protein, which was electrostatically attracted to the matrix. The activated esters then coupled to the *N*-terminus and sole Lys of Grb2 SH3C, leaving the protein covalently attached to the dextran surface. The small molecules were then screened for binding to Grb2 SH3C at 50 μM and 100 μM concentrations (Figure 2-4).

To distinguish binders from non-binders, a practical minimum binding affinity was defined as ≤ 1 mM. The maximal theoretical response observed by SPR depends on the concentration of bound protein and the relative molecular weights of the protein and ligand:

$$R\max_{analyte} = \frac{(MW_{analyte} \times R_{ligand})}{(MW_{ligand})}$$

where ‘analyte’ refers to the small molecule and ‘ligand’ to the SH3C domain. Given 1500 response units (RU) of ligand bound, a protein mass of 7036 Da, and an average small molecule weight of 250 Da, this gives a maximum theoretical response of 53 RU. Thus the minimum desired response level (R_{eq}) could be determined using the formula:

$$R_{eq} = \frac{R\max_{analyte} \times C_{analyte}}{K_d + C_{analyte}}$$

where $C_{analyte}$ represents the concentration of compound, giving a minimum cutoff of 5 RU for the 100 μM concentration.

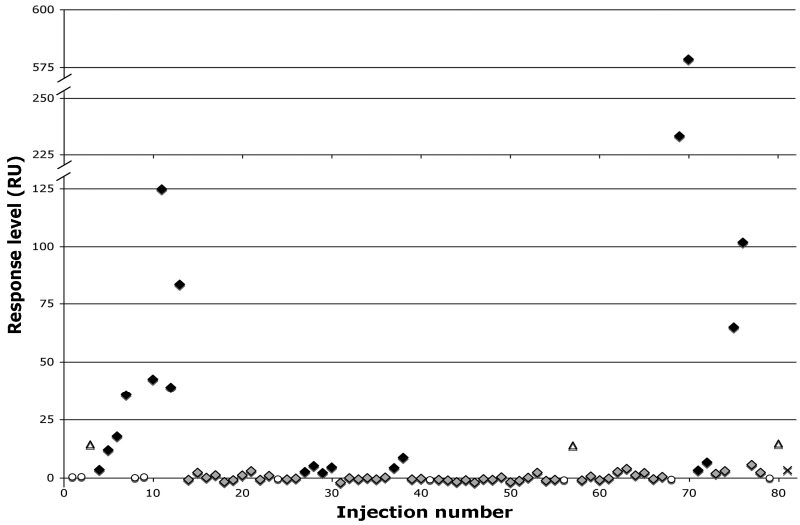


Figure 2-4: Maximum binding response for the 34 molecules at 50 μM and 100 μM concentrations (pairs of adjacent injections). Those molecules exhibiting a response level greater than 5 RU for the 100 μM samples are highlighted (black diamonds) and correspond to compounds: **1**, **3**, **5**, **9**, **41**, **46**, **72**, **147**, **156** and **164**; non-binders are shown as grey diamonds. DMSO-only samples (white circles) and a positive binding control Gab2b peptide (24 μM; white triangles) were injected throughout. A negative control Gab2b peptide double mutant (RxxK → AxxA) is also indicated as the last injection (black cross), showing the lack of significant binding.

Using this threshold, 24 molecules were classified as non-binders (17 with a response level around zero) and ten as binders: **1**, **3**, **5**, **9**, **41**, **46**, **72**, **147**, **156**, **164** (Figure 2-4). Varying levels and quality of interactions were observed by the analysis of their sensorgrams (Figure 2-5; compound **156** was rejected due to poor sensorgram shape). In several cases (molecules **5**, **9**, **147**, **164**) the observed binding behaviour was superstoichiometric. Compound **147**, for instance, displayed superstoichiometric binding, characteristic of undesirable promiscuous interactions,[39] with a response level of nearly 600 RU (over 10 times the maximum theoretical response) and a long dissociation time. This can be explained by the availability of multiple binding sites on Grb2 SH3C or the presence of small-molecule aggregates that have a propensity to adhere to the protein surface. Injections of DMSO-only controls at regular intervals as well as a positive control Gab2b

peptide (24 μM; sequence: IQPPPVNRNLKPDRK-amide) indicated that the integrity of the coupled Grb2 SH3C surface was maintained throughout the experiment (Figure 2-4).

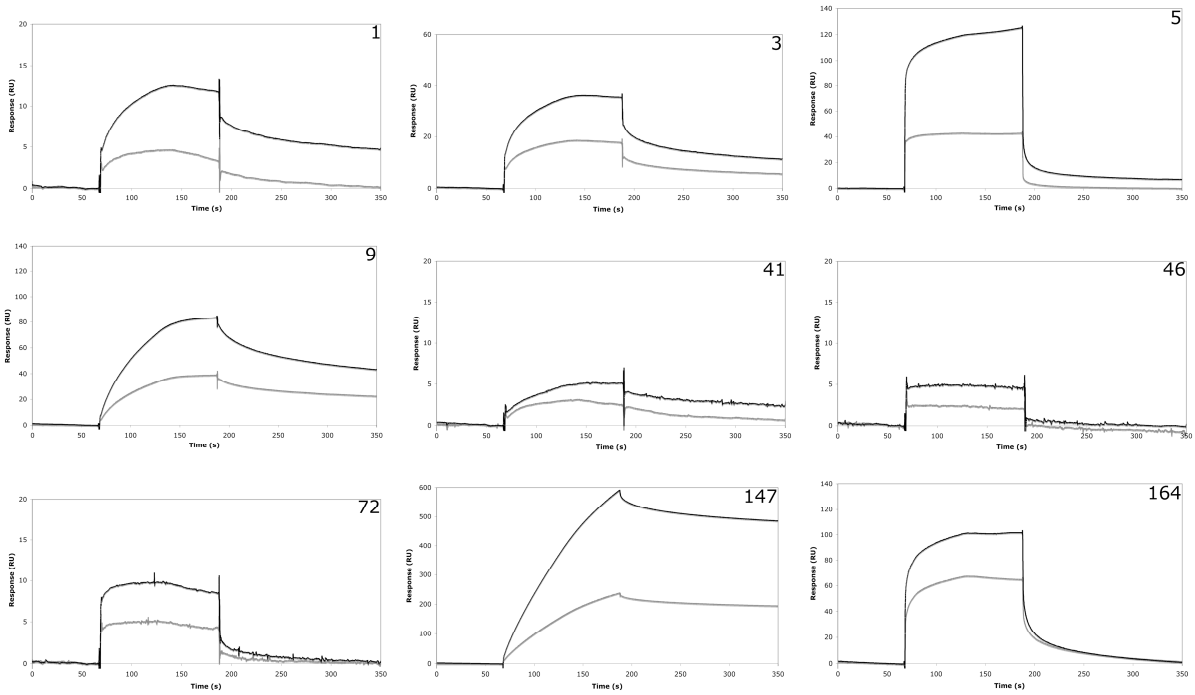


Figure 2-5: SPR sensorgrams of the nine compounds accepted as positive Grb2 SH3C binders. Shown are two traces at different compound concentrations: 50 μM (grey line) and 100 μM (black line).

2.5. Competition binding assay

While the SPR screen reliably measures binding, it gives no information about the binding site. To determine which compounds could competitively target the Gab2 binding groove in Grb2 SH3C, we performed a biochemical competitive binding assay on the full set of 34 molecules. GST–Grb2 SH3C was incubated with each molecule before contact with streptavidin beads coupled to biotinylated Gab2b peptide. Control experiments by Philip Simister indicated that GST alone did not couple to the peptide-bound beads. Neither did GST nor GST–Grb2 SH3C bind to beads lacking peptide, nor to beads coupled to a Gab2 double alanine mutant peptide (RxxK → AxxA) that cannot bind to the Grb2 SH3C domain (Figure 2-6).[23] Thus, interaction with the beads was a direct result of the Gab2b peptide – Grb2 SH3C coupling.

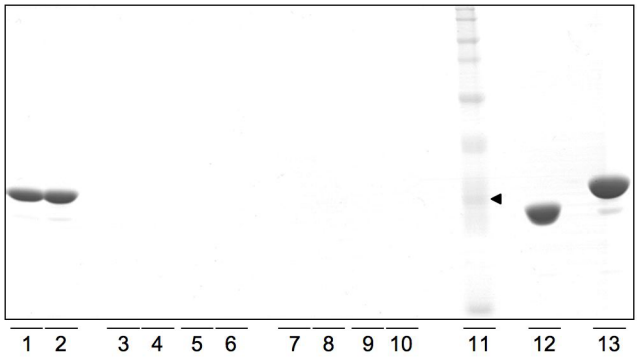


Figure 2-6: 12 % SDS-PAGE gel showing controls for validation of the biochemical competition assay. Samples in gel lanes are: [1, 2] GST–Grb2 SH3C incubated with streptavidin-sepharose beads coupled to Gab2b wild-type (wt) peptide; [3, 4] GST–Grb2 SH3C incubated with streptavidin-sepharose beads coupled to the defunct Gab2b double mutant peptide (*i.e.* RxxK → AxxA); [5, 6] GST–Grb2 SH3C incubated with streptavidin-sepharose beads alone, no peptide coupled; [7, 8] GST alone incubated with streptavidin-sepharose beads coupled to Gab2b wt peptide; [9, 10] GST alone incubated with streptavidin-sepharose beads, no peptide coupled; [11] Protein size markers, a band of 31 kDa is indicated (black arrowhead), which has a size in between that of GST alone (26 kDa), loaded in lane [12], and GST–Grb2 SH3C fusion protein (33 kDa), loaded in greater amount (30 µg) in lane [13] to show the quality of protein used in the assay.

Given this assay's lower sensitivity, a stringent threshold of 20 % minimum inhibition was applied to eliminate ambiguous-, weak-, or non-binders. This represents a margin greater than the error on any individual sample, and five times the average error on the data points (Figure 2-7a). Consequently, 24 molecules showed no binding (Figure 2-7B, grey bars), and ten molecules: **3**, **5**, **41**, **46**, **50**, **55**, **72**, **139**, **147**, **164**, exhibited various levels of competitive inhibition > 20 % (Figure 2-7b, black bars). Importantly, seven of these competitive inhibitors overlapped with those determined to bind well by SPR: **3**, **5**, **41**, **46**, **72**, **147**, **164**. However, three of these molecules: **50**, **72**, **164**, displaying nearly complete inhibition precipitated during the assay. This made them problematic to handle and may have contributed to their apparent superior inhibitory behaviour (compound **72** was the least soluble); molecule **46** also displayed poor aqueous solubility, and so these four compounds were rejected. The superstoichiometric binder from SPR analysis – compound **147** – was excluded too, leaving a subset of three compounds (**3**, **5**, **41**) for further assessment. Above concentrations of 200 µM, these compounds displayed superstoichiometric binding by SPR, preventing the collection of meaningful affinity data using that method. Instead, the

biochemical competition assay was employed using increasing compound concentrations, with the aim of assessing the affinities of these compounds for Grb2 SH3C.

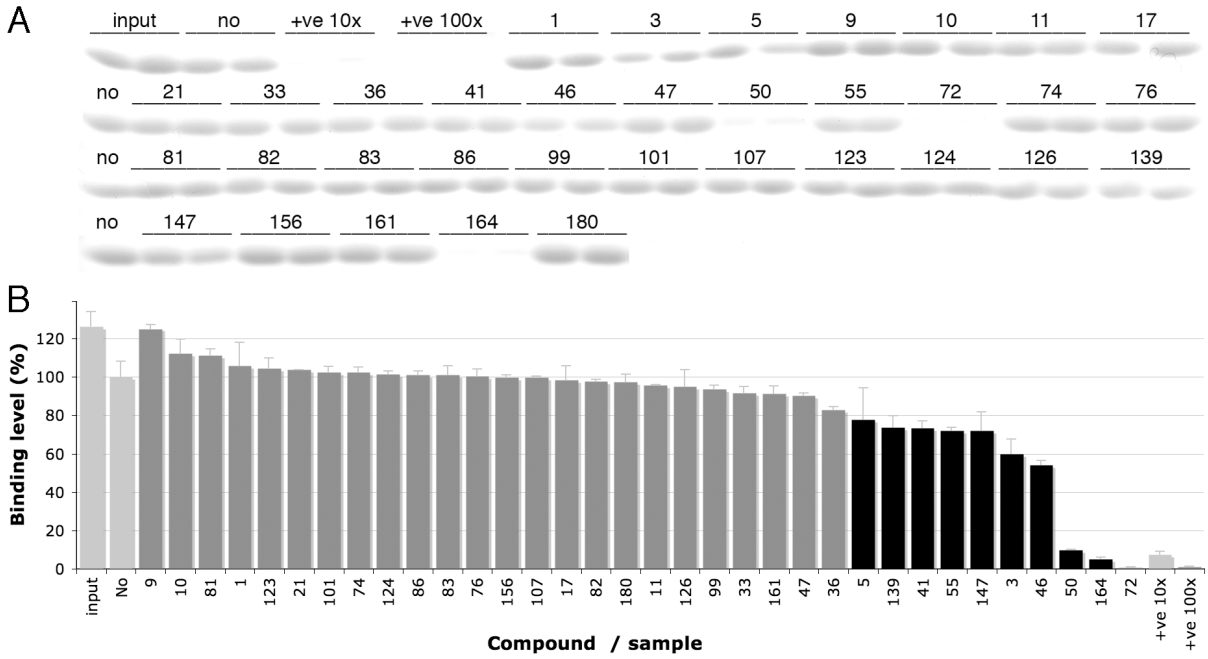


Figure 2-7: Grb2 SH3C competitive binding screen with 34 hit molecules using a biochemical competition assay. Purified GST–Grb2 SH3C was incubated with each compound (5 mM) in duplicate before contact with streptavidin-sepharose beads previously coupled to biotinylated Gab2b peptide (Bio-EAHK-IQPPPVNRNLKPDRK-amide). Washed beads were run on a 12 % SDS–polyacrylamide gel and the extent of binding visualized by SDS–PAGE and Coomassie blue dye staining (A), and the band intensities quantified by densitometry (B). In (A), 'no' refers to no compound—the DMSO-only control. Numbering indicates the compound used, and relates to the rank of each in the virtual screen. Unbiotinylated Gab2b peptide was used as a positive binding control for inhibition (+ve) at two concentrations. In (B), chart bars are shaded as follows: the ten positive binders showing > 20 % inhibition (black); non-binders (grey); controls, or 10 μg of GST–Grb2 SH3C input (light grey). Error bars indicate the spread of the data.

This assay, like other classical biochemical assays, relies on a simple kinetic model from the law of mass action. Under this assumption, a binding equilibrium exists between ligand—in this case, bead-bound Gab2b peptide—and Grb2 SH3C. The fraction of binding sites occupied depends on the dissociation constant K_d of the ligand and the amount of ligand present:

$$Occupancy = \frac{[L]}{K_d + [L]}$$

The addition of a competitive inhibitor disrupts this binding, reducing the amount of ligand bound. At low concentrations of inhibitor the ligand is fully bound, and at high concentrations of inhibitor only nonspecifically-bound ligand remains. The difference between these values is the specific binding of the ligand, and the concentration of inhibitor that reduces specific biding by 50 % is the IC_{50}. The curve is mathematically defined as:

$$[LR] = NS + \frac{total - NS}{1 + 10^{\log[I] - \log IC_{50}}}$$

Where [*LR*] is the concentration of the protein-peptide complex, *total* is the maximum observed binding, *NS* is the nonspecific binding, and [*I*] is the concentration of inhibitor. If the ligand and inhibitor compete for a single binding site, an 81-fold increase in inhibitor concentration reduces the specific binding from 90 % to 10 %, and the great majority of the curve fits within a two log unit range.[40]

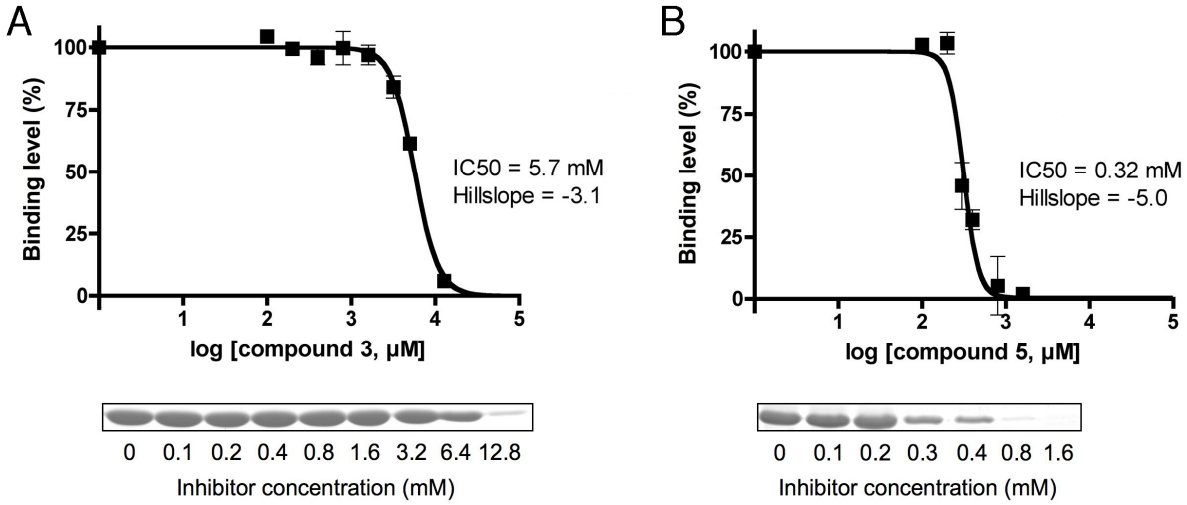


Figure 2-8: Concentration-response curves determined by the biochemical competition assay for two competitively binding compounds. Increasing concentrations of (A) compound **3**, and (B) compound **5**, were incubated with GST–Grb2 SH3C before mixing with streptavidin beads coupled to a biotinylated Gab2b peptide. (A) and (B) lower panels: Coomassie blue-dyed bands of the GST–Grb2 SH3C/beads samples run on a 12 % SDS polyacrylamide gel; upper panels: densitometric quantification and curve-fit of the protein band intensities. The calculated IC_{50} values and Hill slope parameters are shown. Error bars show standard deviations; the gels are representative of four to six repeats.

Using this competition assay, compound **41** demonstrated unquantifiably weak inhibition in the limited dose range used (up to 20 mM). Compound **3** yielded a concentration-response (inhibition) curve with an apparent IC_{50} of 5.7 mM, indicating low affinity (Figure 2-8A). Despite its structural similarity to compound **3**, compound **5** bound competitively to Grb2 SH3C with an apparent IC_{50} of 320 μM, representing an 18-fold increase (Figure 2-8B). Notwithstanding this positive competitive inhibitory effect, an important feature of the inhibition curves for both compounds **3** and **5** is their steepness, as complete inhibition takes place over only one log unit. This represents markedly non-stoichiometric binding, which is quantified as the Hill slope *n*:

$$[LR] = \frac{[LR]_{\max} \times [L]^n}{K_d + [L]^n}$$

For simple one-to-one binding, the Hill slope is -1.0, while in this instance both molecules display very high Hill slopes (Figure 2-8). Molecules displaying such binding characteristics are less ideal but not purely artifactual.[41]

2.6. Calculation of tautomeric forms of the dihydro-*s*-triazine motif

The dihydro-*s*-triazine motif present in seven of the compounds (**1**, **3**, **5**, **9**, **11**, **41**, **47**) exists as the mono HCl salt. Either of the basic nitrogen atoms on the ring could be protonated, so the relative thermodynamic stability of the two annular tautomers was assessed by quantum chemical calculations. Using the composite CBS-QB3 method,[42] the heterocycle is calculated to be protonated on the 5-position (Figure 2-9a). This is consistent with neutron diffraction experiments on a related molecule, cycloguanil•HCl (Figure 2-9), which was found to be protonated on the same nitrogen.[43]

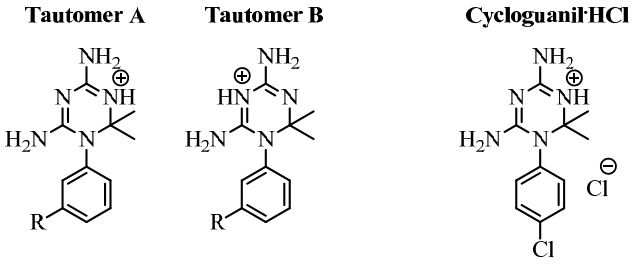


Figure 2-9: Two annular tautomers of the dihydro-*s*-triazine motif. The relative energies (where R = H) were calculated in the gas phase using the composite quantum chemical method CBS-QB3. The calculated free energy of tautomer A is 12.94 kcal.mol^{-1} lower than that of tautomer B. This is consistent with neutron diffraction measurements of cycloguanil•HCl.

2.7. Discussion

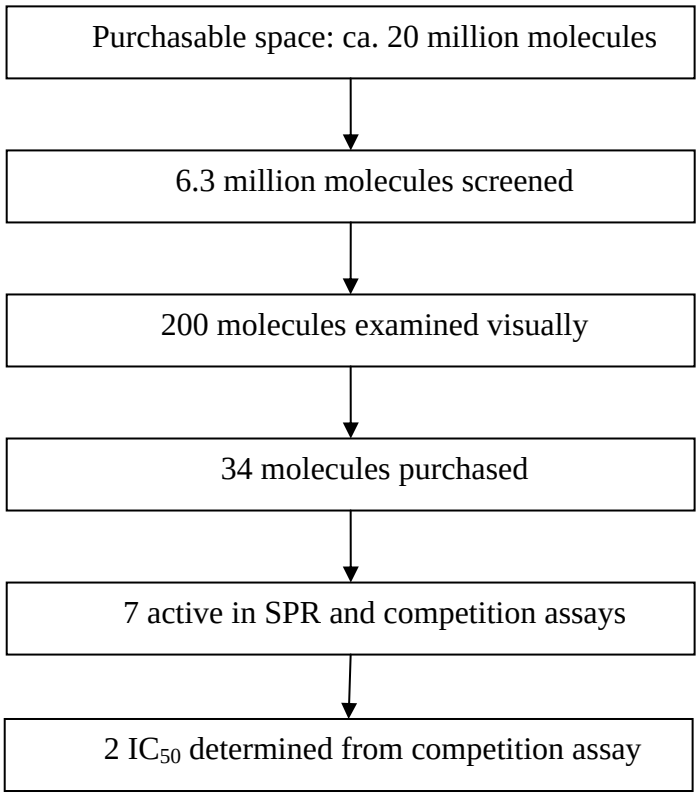


Figure 2-10: Virtual screening process from purchasable chemical space to two binders of to Grb2 SH3C.

Among the 34 molecules purchased, seven (20 %) displayed activity in two orthogonal assays (Figure 2-10). While a proportion of the binding observed may be a result of multiple interaction sites on Grb2 SH3C, the biochemical competition assay confirms two molecules (compounds **3** and **5**) that interact competitively with the Gab2b peptide-binding surface of Grb2 SH3C. Their related structures provide some structure-activity relationship (Figure

2-2); in fact, among the 34 molecules, seven, **1**, **3**, **5**, **9**, **11**, **41**, **47**, featured a common dihydro-*s*-triazine motif which has previously been investigated for the inhibition of dihydrofolic reductase[44] and voltage-gated sodium channels.[45] These seven molecules span several orders of magnitude in binding affinity, ranging from 320 μM to undetectably low levels. Consistent with previous peptide array studies,[23] only mixed cationic and hydrophobic molecules bound to Grb2 SH3C, whereas molecules **11** and **47**, which are cationic but lack significant hydrophobic motifs, showed no activity in either assay. Compound **9**, which differs from **3** in having a hydrophilic polyethylene glycol linker in place of a hydrophobic alkyl chain, similarly shows no activity. The most strongly binding compounds **3** and **5** have linkers of equal length, though the more rigid alkene linker for compound **5** may contribute to its improved binding affinity by lessening the entropic penalty of binding in an extended conformation.

2.8. Conclusion and future work

This is the first computational screen reported for the identification of small molecules able to bind to Grb2 SH3C. Given the challenging nature of targeting protein-protein interaction surfaces, it is encouraging that a virtual screen allowed identification of small molecule *in vitro* binders. Despite the modest number of compounds acquired for *in vitro* testing, it is remarkable that over 20 % of these displayed inhibitory activity in two orthogonal screens. The identified molecules, however, are very weak, non-stoichiometric binders.

It remains to be explored whether the potency of these molecules can be further developed, potentially allowing cell-based studies. The modest affinities achieved from these commercially available molecules (*cf.* 15-mer Gab2b peptide, K_d 3.2 μM towards Grb2 SH3C) and their likely multi-point interaction on the SH3 surface suggests that a more

bottom-up approach could be beneficial for producing effective inhibitors. This will be explored in Chapters 3 & 4. The virtual screen is not expected to rank the compounds perfectly, so visual inspection of molecules beyond the top 200 may identify other interesting compounds and greater diversity in predicted binding poses. Additionally, more molecules could be acquired from among the highest-ranked binders and tested *in vitro*. This process would be aided by the SAR generated from the dihidro-*s*-triazine molecules, which suggests that small or uncharged compounds are unlikely to be true binders.

Wider virtual screening may probe chemical space that is more favourable for protein-protein interaction inhibitors, though delineating that space is an ongoing challenge.[46,47] It is possible that the initial choice of library, containing "Lipinski-compliant" molecules, was itself biased away from more fruitful areas of chemical space. With additional computational resources, the screening library could be expanded to the larger set of purchasable molecules. This might identify larger or more heteroatom-rich molecules that may make larger contacts with the surface of Grb2 SH3C. Additionally, numerous other scoring functions exist including an "extra precision" mode in GLIDE.[48] A subset of molecules which scored highly in the present screen (perhaps ca. 100,000) could be docked again using these more precise scoring functions, potentially identifying stronger binding interactions and creating a more accurate ranking of the compounds. *In vitro* assessment of these new compounds could help further delineate the structural features required to bind to Grb2 SH3C.

2.9. Experimental details

2.9.1. SPR binding screen

SPR was performed using a Biacore T200 optical biosensor. A fresh CM5 chip was docked to the instrument and hydrated with duplicate 12 s injections of 50 mM NaOH, 10

mM Glycine•HCl pH 1.5, and 0.1 % SDS at 50 μL.min^{-1} flow rate. The flow cells were then normalized with 70 % glycerol using the default normalization wizard. The chip was primed into HBS-N running buffer (10 mM Hepes pH 7.4, 150 mM NaCl), and the flow cells activated with a 1:1 mixture of 0.4 M EDC and 0.1 M NHS for 10 minutes at 10 μL.min^{-1} and 25 ˚C. Untagged Grb2 SH3C (100 μM in pH 4.0 acetate buffer) was coupled to the surface for 8 minutes to an immobilization level of 1500 RU. After coupling, all flow cell surfaces were deactivated with 4 x 30 s injections of a 1:1 mixture of 1.0 M ethanolamine and running buffer. The chip was subsequently primed twice into the experimental running buffer (HBS-N, 5.0 % DMSO, 0.01 % Tween 20).

Each molecule was dissolved in running buffer (HBS-N, 5.0 % DMSO, 0.01 % Tween 20) at 50 μM and 100 μM concentrations and then screened using 120 s contact time at 40 μL.min^{-1} flow rate and 25 ˚C, with a data collection rate of 10 Hz. The data were analyzed using standard double referencing procedures by subtracting the response from the protein-free cell and the response from a blank injection of running buffer from the response from the SH3C flow cell.[49]

2.9.2. Competition binding assay

GST–Grb2 SH3C (10 μg per sample) was pre-incubated with each inhibitor (5 mM, initially) in duplicate for 2 hours at 4 ˚C in assay buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT, 0.05 % Tween 20, 5 % DMSO). Biotinylated Gab2b peptide (binding affinity 3.2 μM for Grb2) was synthesized as previously described.[23] 0.3 nmol of peptide per μL of packed streptavadin-sepharose beads were incubated together for 30 minutes in assay buffer before extensive washing of the beads to remove any unbound peptide. Next, 8 μL of peptide-coated beads were transferred to each assay tube containing GST–Grb2 SH3C and compound for 30 minutes, the solution clarified by centrifugation (2000 x *g* for 1 minute) and the beads washed three times with PBS-T (phosphate-buffered saline, 0.05 %

Tween 20). The samples were analyzed by SDS-PAGE and visualized after staining the protein bands with Coomassie Blue dye (Figure 2-7A). As controls, parallel samples were run in the absence of inhibitor, and with preincubation of 10 or 100 x molar excess of free, unbiotinylated Gab2b peptide. The band intensities, indicating the levels of GST–Grb2 SH3C binding to the beads, were further quantified by densitometry and scaled relative to the control sample without inhibitor (100 % level; Figure 2-7). This normalization allowed the cross-comparison of multiple gels.

For concentration-response experiments, increasing concentrations of compound were used. The data were fitted to a sigmoidal inhibition model using nonlinear regression analysis in GraphPad Prism software (version 4.00 GraphPad Software, San Diego California USA).

2.9.3. Spectroscopic details of compounds **3** *and* **5**

1-(3-(4-(3-Aminophenyl)butyl)phenyl)-6,6-dimethyl-1,6-dihydro-1,3,5-triazine-2,4-diamine•HCl (3)

δ_H (500 MHz, $(CD_3)_2SO$) 9.00 (1H, br s), 7.64 (1H, br s), 7.44 (1H, t, *J* 8.0), 7.35 (2H, br s), 7.34 (1H, d, *J* 7.8), 7.19-7.15 (2H, m), 6.89 (1H, t, *J* 7.6), 6.38-6.34 (2H, m), 6.31 (1H, d, *J* 7.4), 6.25 (1H, br s), 4.91 (2H, s), 2.66 (2H, t, *J* 7.3), 2.44 (2H, t, *J* 8.0), 1.66-1.51 (4H, m), 1.10 (6H, d, *J* 2.1); δ_C (125 MHz, $(CD_3)_2SO$) 157.6, 157.2, 148.5, 144.5, 144.5, 142.6, 134.7, 129.9, 129.7, 128.6, 127.1, 115.9, 113.9, 111.5, 69.6, 35.2, 34.5, 30.3, 30.3, 27.3, 27.2; HRMS (ESI) found 365.2442, $C_{21}H_{29}N_6^+$ $[M+H]^+$ requires 365.2448.

6,6-Dimethyl-1-(3-((1*E*,3*E*)-4-(3-nitrophenyl)buta-1,3-dien-1-yl)phenyl)-1,6-dihydro-1,3,5-triazine-2,4-diamine•HCl (5)

δ_H (500 MHz, $(CD_3)_2SO$) 9.04 (1H, s), 8.35 (1H, t, *J* 1.8), 8.10 (1H, dd, *J* 8.2, 0.6), 8.01 (1H, d, *J* 7.9), 7.71-7.64 (2H, m), 7.57 (1H, s), 7.53 (1H, t, *J* 7.8), 7.39-7.32 (1H, m), 7.35

(4H, br s), 7.31-7.24 (2H, m), 6.94 (1H, d, *J* 5.4), 6.90 (1H, d, *J* 5.4), 6.42 (1H, br s), 1.39 (6H, d, *J* 7.5); δ_C (125 MHz, $(CD_3)_2SO$) 157.6, 157.2, 148.4, 138.9, 138.8, 135.4, 133.4, 132.5, 132.1, 131.0, 130.5, 130.3, 130.2, 129.2, 127.8, 127.8, 122.1, 120.6, 69.6, 27.3, 27.3; HRMS (ESI) found 391.1872, $C_{21}H_{23}N_6O_2^+$ $[M+H]^+$ requires 391.1877.

2.9.4. Ab initio *structure calculations*

The dihydro-s-triazine was built using Gaussview (version 5)[50] and protonated on each of the two basic ring positions (Figure 2-9). The two annular tautomers were minimized in the gas phase using the composite quantum chemical method CBS-QB3,[42] using the default parameters in Gaussian09 (revision B.01).[51] Based on the CBS-QB3 computed free energies, tautomer A is expected to be the more stable form of the molecule by 12.94 kcal.mol^{-1} in the gas phase at 298.16 K (Figure 2-11).

Tautomer A
CBS-QB3 (0 K) = -701.163938 CBS-QB3 Energy = -701.148641
CBS-QB3 Enthalpy = -701.147697 CBS-QB3 Free Energy = -701.205483

Tautomer B
CBS-QB3 (0 K) = -701.143427 CBS-QB3 Energy = -701.128176
CBS-QB3 Enthalpy = -701.127232 CBS-QB3 Free Energy = -701.184869

Figure 2-11: Thermodynamic properties of tautomers A and B calculated using the CBS-QB3 composite method. All energies are listed in Hartrees at 298.16 K. The calculated free energy of tautomer A is 12.94 kcal.mol^{-1} lower than that of tautomer B.

2.10. References

(1) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. Identification and Specificity Studies of Small-Molecule Ligands for SH3 Protein Domains. *J. Med. Chem.* **2004**, *47*, 5405–5417.

(2) Inglis, S.; Jones, R.; Fritz, D.; Stojkoski, C.; Booker, G.; Pyke, S. Synthesis of 5-, 6- and 7-substituted-2-aminoquinolines as SH3 domain ligands. *Org. Biomol. Chem.* **2005**, *3*, 2543–2557.

(3) Inglis, S. R.; Jones, R. K.; Booker, G. W.; Pyke, S. M. Synthesis of N-benzylated-2-aminoquinolines as ligands for the Tec SH3 domain. *Bioorg. Med. Chem. Let.* **2006**, *16*, 387–390.

(4) Smith, J. A.; Jones, R. K.; Booker, G. W.; Pyke, S. M. Sequential and Selective Buchwald−Hartwig Amination Reactions for the Controlled Functionalization of 6-

Bromo-2-chloroquinoline: Synthesis of Ligands for the Tec Src Homology 3 Domain. *J. Org. Chem.* **2008**, *73*, 8880–8892.
(5) Atatreh, N.; Stojkoski, C.; Smith, P.; Booker, G. W.; Dive, C.; Frenkel, A. D.; Freeman, S.; Bryce, R. A. In silico screening and biological evaluation of inhibitors of Src-SH3 domain interaction with a proline-rich ligand. *Bioorg. Med. Chem. Let.* **2008**, *18*, 1217–1222.
(6) Sharma, S. V.; Oneyama, C.; Yamashita, Y.; Nakano, H.; Sugawara, K.; Hamada, M.; Kosaka, N.; Tamaoki, T. UCS15A, a non-kinase inhibitor of Src signal transduction. *Oncogene* **2001**, *20*, 2068–2079.
(7) Oneyama, C.; Nakano, H.; Sharma, S. V. UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug. *Oncogene* **2002**, *21*, 2037–2050.
(8) Oneyama, C.; Agatsuma, T.; Kanda, Y.; Nakano, H.; Sharma, S. V.; Nakano, S.; Narazaki, F.; Tatsuta, K. Synthetic Inhibitors of Proline-Rich Ligand-Mediated Protein-Protein Interaction: Potent Analogs of UCS15A. *Chem. Biol.* **2003**, *10*, 443–451.
(9) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.
(10) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Heroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17889–17894.
(11) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417–3423.
(12) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009**, *394*, 747–763.
(13) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6*, 3850–3856.
(14) Knox, A. J. S.; Meegan, M. J.; Carta, G.; Lloyd, D. G. Considerations in Compound Database Preparation“Hidden” Impact on Virtual Screening Results. *J. Chem. Inf. Model.* **2005**, *45*, 1908–1919.
(15) Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A. The Effect of Ligand-Based Tautomer and Protomer Prediction on Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 2742–2748.
(16) Chen, I.-J.; Foloppe, N. Conformational Sampling of Druglike Molecules with MOE and Catalyst: Implications for Pharmacophore Modeling and Virtual Screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773–1791.
(17) Mobley, D. L. Let’s get honest about sampling. *J Comput. Aided Mol. Des.* **2012**, *26*, 93–95.
(18) Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.
(19) Lu, N.; Kofke, D. A.; Woolf, T. B. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *J. Comput. Chem.* **2004**, *25*, 28–40.
(20) Coupez, B.; Lewis, R. A. Docking and scoring--theoretically easy, practically impossible? *Curr. Med. Chem.* **2006**, *13*, 2995–3003.

(21) Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J. Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications. *J. Med. Chem.* **2010**, *53*, 8461–8467.
(22) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52*, 867–881.
(23) Harkiolaki, M.; Tsirka, T.; Lewitzky, M.; Simister, P. C.; Joshi, D.; Bird, L. E.; Jones, E. Y.; O'Reilly, N.; Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the SH3C Domain of Grb2. *Structure* **2009**, *17*, 809–822.
(24) Irwin, J. J.; Shoichet, B. K. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(25) 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.* **2012**, *52*, 1757–1768.
(26) 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 Del. Res.* **2001**, *46*, 3–26.
(27) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(28) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
(29) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(30) Sabban, S. *Development of an in vitro model system for studying the interaction of Equus caballus IgE with its high-affinity FcεRI receptor.*; Doctoral dissertation, University of Sheffield; Doctoral dissertation, University of Sheffield, 2011.
(31) Homola, J.; Yee, S. S.; Gauglitz, G. Surface plasmon resonance sensors: review. *Sensors and Actuators B: Chemical* **1999**, *54*, 3–15.
(32) Pattnaik, P. Surface plasmon resonance. *Appl. Biochem. Biotechnol.* **2005**, *126*, 79–92.
(33) Schasfoort, R. B. .; Tudos, A. J. *Handbook Of Surface Plasmon Resonance*; Royal Society of Chemistry, 2008.
(34) Giannetti, A. M. Chapter Eight - From Experimental Design to Validated Hits: A Comprehensive Walk-Through of Fragment Lead Identification Using Surface Plasmon Resonance. In *Methods in Enzymology*; Lawrence C. Kuo, Ed.; Academic Press, 2011; Vol. Volume 493, pp. 169–218.
(35) Day, Y. S. N.; Baird, C. L.; Rich, R. L.; Myszka, D. G. Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. *Protein Sci.* **2002**, *11*, 1017–1025.
(36) Cannon, M. J.; Papalia, G. A.; Navratilova, I.; Fisher, R. J.; Roberts, L. R.; Worthy, K. M.; Stephen, A. G.; Marchesini, G. R.; Collins, E. J.; Casper, D.; Qiu, H.; Satpaev, D.; Liparoto, S. F.; Rice, D. A.; Gorshkova, I. I.; Darling, R. J.; Bennett, D. B.; Sekar, M.; Hommema, E.; Liang, A. M.; Day, E. S.; Inman, J.; Karlicek, S. M.; Ullrich, S. J.; Hodges, D.; Chu, T.; Sullivan, E.; Simpson, J.; Rafique, A.; Luginbühl, B.; Westin, S. N.; Bynum, M.; Cachia, P.; Li, Y.-J.; Kao, D.; Neurauter, A.; Wong, M.;

Swanson, M.; Myszka, D. G. Comparative analyses of a small molecule/enzyme interaction by multiple users of Biacore technology. *Anal. Biochem.* **2004**, *330*, 98–113.

(37) Papalia, G. A.; Leavitt, S.; Bynum, M. A.; Katsamba, P. S.; Wilton, R.; Qiu, H.; Steukers, M.; Wang, S.; Bindu, L.; Phogat, S.; Giannetti, A. M.; Ryan, T. E.; Pudlak, V. A.; Matusiewicz, K.; Michelson, K. M.; Nowakowski, A.; Pham-Baginski, A.; Brooks, J.; Tieman, B. C.; Bruce, B. D.; Vaughn, M.; Baksh, M.; Cho, Y. H.; Wit, M. D.; Smets, A.; Vandersmissen, J.; Michiels, L.; Myszka, D. G. Comparative analysis of 10 small molecules binding to carbonic anhydrase II by different investigators using Biacore technology. *Anal. Biochem.* **2006**, *359*, 94–105.

(38) Navratilova, I.; Papalia, G. A.; Rich, R. L.; Bedinger, D.; Brophy, S.; Condon, B.; Deng, T.; Emerick, A. W.; Guan, H.-W.; Hayden, T.; Heutmekers, T.; Hoorelbeke, B.; McCroskey, M. C.; Murphy, M. M.; Nakagawa, T.; Parmeggiani, F.; Qin, X.; Rebe, S.; Tomasevic, N.; Tsang, T.; Waddell, M. B.; Zhang, F. F.; Leavitt, S.; Myszka, D. G. Thermodynamic benchmark study using Biacore technology. *Anal. Biochem.* **2007**, *364*, 67–77.

(39) Giannetti, A. M.; Koch, B. D.; Browner, M. F. Surface Plasmon Resonance Based Assay for the Detection and Characterization of Promiscuous Inhibitors. *J. Med. Chem.* **2008**, *51*, 574–580.

(40) *Comprehensive Toxicology*; Newnes, 2010.

(41) Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **2006**, *49*, 7274–7277.

(42) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.

(43) Schwalbe, C. H.; Williams, G. J.; Koetzle, T. F. Structure of cycloguanil hydrochloride by neutron diffraction. *Acta Crystallogr. C* **1989**, *45 (Pt 3)*, 468–471.

(44) Baker, B. R.; Lourens, G. J. Irreversible enzyme inhibitors. CIX. Candidate irreversible inhibitors of dihydrofolic reductase derived from 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-phenyl-s-triazine. 3. *J. Med. Chem.* **1968**, *11*, 26–33.

(45) Ma, X.; Poon, T.-Y.; Wong, P. T. H.; Chui, W.-K. Synthesis and in vitro evaluation of 2,4-diamino-1,3,5-triazine derivatives as neuronal voltage-gated sodium channel blockers. *Bioorg. Med. Chem. Let.* **2009**, *19*, 5644–5647.

(46) Sperandio, O.; Reynès, C. H.; Camproux, A.-C.; Villoutreix, B. O. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today* **2010**, *15*, 220–229.

(47) Villoutreix, B. O.; Labbé, C. M.; Lagorce, D.; Laconde, G.; Sperandio, O. A leap into the chemical space of protein-protein interaction inhibitors. *Curr. Pharm. Des.* **2012**, *18*, 4648–4667.

(48) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(49) Myszka, D. G. Improving biosensor analysis. *J. Mol. Recognit.* **1999**, *12*, 279–284.

(50) Dennington, R.; Keith, T.; Millam, J. *GaussView*; Semichem Inc.: Shawnee Mission KS, 2009.

(51) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda,

Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. Gaussian 09, Revision B.01 **2009**.

3. Benzoylurea-based PPII peptidomimetics

3.1. Introduction

The polyproline type II helix (PPII) is a structural motif critical to a range of cell-signalling processes (*vide supra* Chapter 1.2.3). Compared to the common α-helix, PPII helices are more extended (ca. 9 Å per turn vs. 5.4 Å) and feature a three-fold axis of symmetry (Figure 3-1). Proline's unique structural features promote PPII formation,[1] although recent work has emphasized that many other peptide sequences, including polyalanine[2] and polylysine[3] form a PPII structure. Solvent polarity is important to PPII helix formation,[4] though intramolecular salt bridges do not appear to be stabilizing.[5] In aqueous solution the PPII helix is rigid, so upon binding to a partner there is only a small loss in entropy, making PPII well-adapted for participating in PPIs.[6] While proline-rich sequences are ubiquitous (> 100 *E. coli* proteins have at least three consecutive Pro), ribosomal synthesis of these proteins is challenging. Indeed, *in vivo* protein elongation stalls for even short Pro (or ProGly) sequences, requiring a ubiquitously-conserved enzyme EF-P to continue synthesis.[7,8]

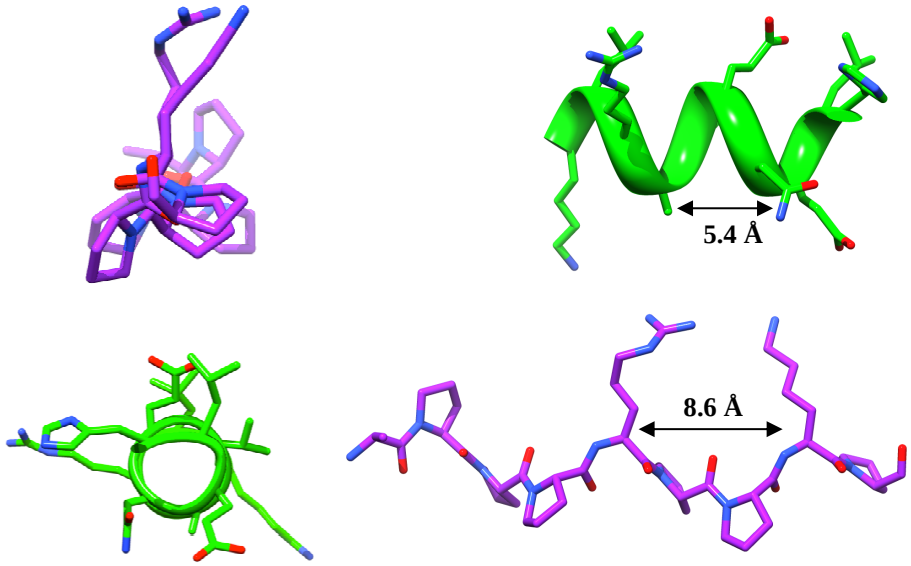


Figure 3-1: Comparison of α-helix (green) and PPII helix (purple) structures. When viewed parallel to the helix axis, the PPII helix has a 3-fold axis of symmetry (top left), compared to the unsymmetrirical α helix (bottom left). The PPII helix is also significantly longer, with one turn of the PPII helix corresponding to approximately 1.5 turns of the α helix (right). The PPII sequence here is the Gab2a peptide (APPPRPPKP).

Few synthetic PPII mimics have been developed. One approach is not to mimic the helix, but rather to functionalize a poly-Pro sequence on the pyrrolidine rings to exploit the natural helicity of these sequences (Figure 3-2a). Such "proline-templated amino acids" have been used in the design of cell-penetrating peptides[9] and in investigating the binding site of kinases.[10] Dipeptide Pro-Pro mimics have been designed which covalently constrain adjacent residues to adopt a PPII conformation.[11] A peptide incorporating such a mimic was found to bind to the Fyn SH3 domain by NMR and ITC, but with an affinity of 62 μM vs. 18 μM for the unconstrained peptide (Figure 3-2b). A different bicyclo mimic from the Geyer group mimics the Ser-Pro dipeptide unit, with a hexapeptide-mimic derived from three such units forming an ordered PPII structure by NMR (Figure 3-2c).[12] For tripeptides, particular effort has been directed towards the Pro-Leu-Gly-NH_2 neuropeptide, which modifies dopamine expression in the brain. Structural evidence suggests that this sequence binds either as a β-turn or as a PPII helix, so mimics have been designed in both conformations to investigate the selectivity of binding. PPII mimics were found to be

positive allosteric modulators of dopamine receptors, suggesting that this conformation may be critical *in vivo* (Figure 3-2d).[13,14] Additionally, a wholly synthetic 3-azabicyclo[3.1.0]hexane was shown to mimic the PPII structure using solution NMR experiments (Figure 3-2e), but biological data is lacking.[15] Despite these successes, these molecules incorporate significant peptidic elements, so the design of an entirely synthetic PPII peptidomimetic is appealing.

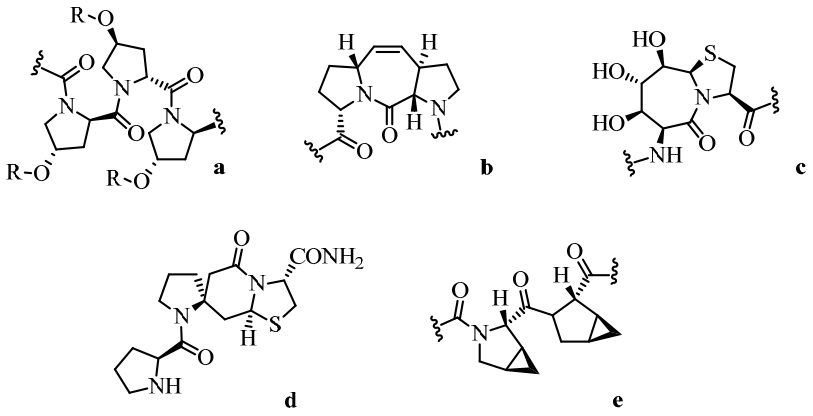


Figure 3-2: PPII helix mimics: (a) a Pro-templated amino acid, (b) Pro-Pro dipeptide mimic (c) Ser-Pro dipeptide mimic, (d) bicyclic Pro-Leu-Gly mimic, (e) 3-azabicyclo[3.1.0]hexane scaffold.

3.2. Peptidomimetic design

The key challenge in mimicking PPII helices is the development of a scaffold which matches the extended length of these peptides. While existing scaffolds have not been designed for PPII mimicry, if one fortuitously matched the structure it would simplify synthesis, so previous Hamilton group mimics were computationally modelled for similarity to the Gab2a PPII peptide. An adapted benzoylurea scaffold,[16–18] previously used in the group to mimic α-helices, was found to closely match the desired conformation. Two aromatic rings and one urea linkage mimic three turns of an α-helix, including *i*, *i*+4, and *i*+7 residues. Instead, given the ca. 50% greater length of the PPII helix, the same scaffold might match the *i*, *i*+1/*i*+2, and *i*+3 residues of a PPII helix.

Solution-phase NMR[16] and X-ray crystallography[17] studies indicate that the central benzoylurea moiety is planar due to an intramolecular hydrogen bond. Further modelling to the PM3 (semi-empirical) level of theory[19] calculates that the central urea moiety remains planar even with butyl chains on the benzene rings. Additionally the calculated C_β-C_β distance measured from the *ortho*-positions of the scaffold could be 8.95 Å (Figure 3-3), comparing favourably to the crystallographically-determined distance of 8.56 Å for the Arg and Lys residues of Gab2a.

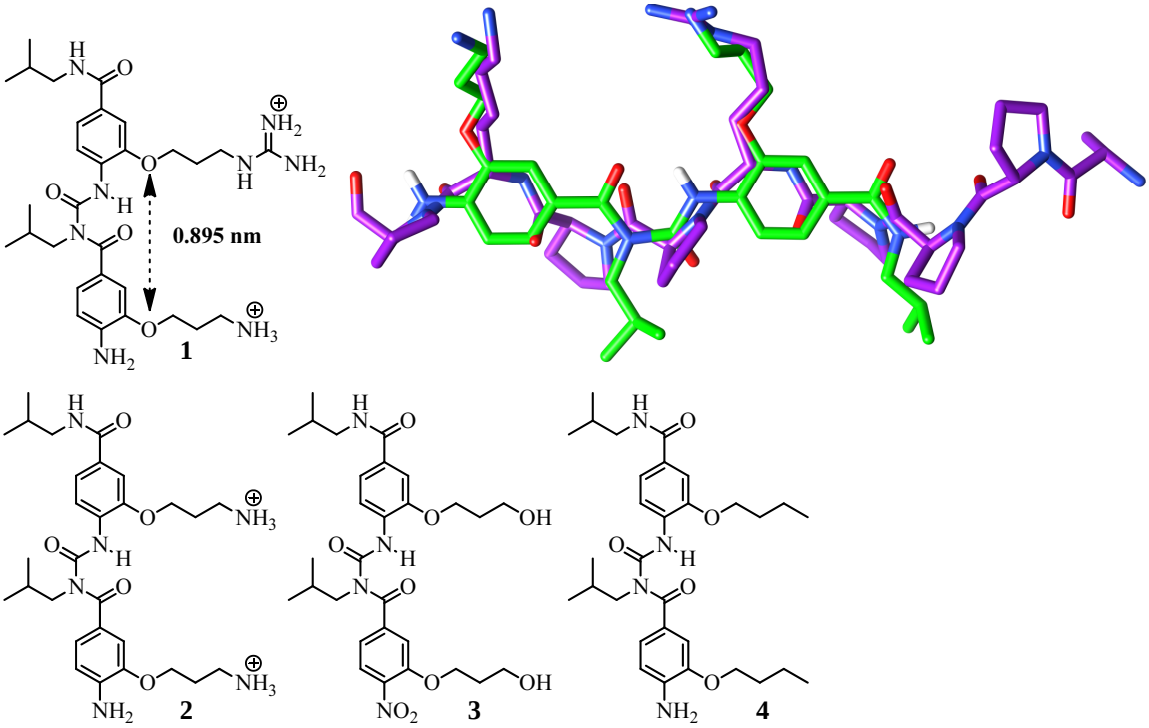


Figure 3-3: Clockwise from top left: Structure of mimetic **1**; overlay of mimetic **2** (green) and Gab2a peptide (purple); control mimetic **4**; control mimetic **3**; simplified mimetic **2**.

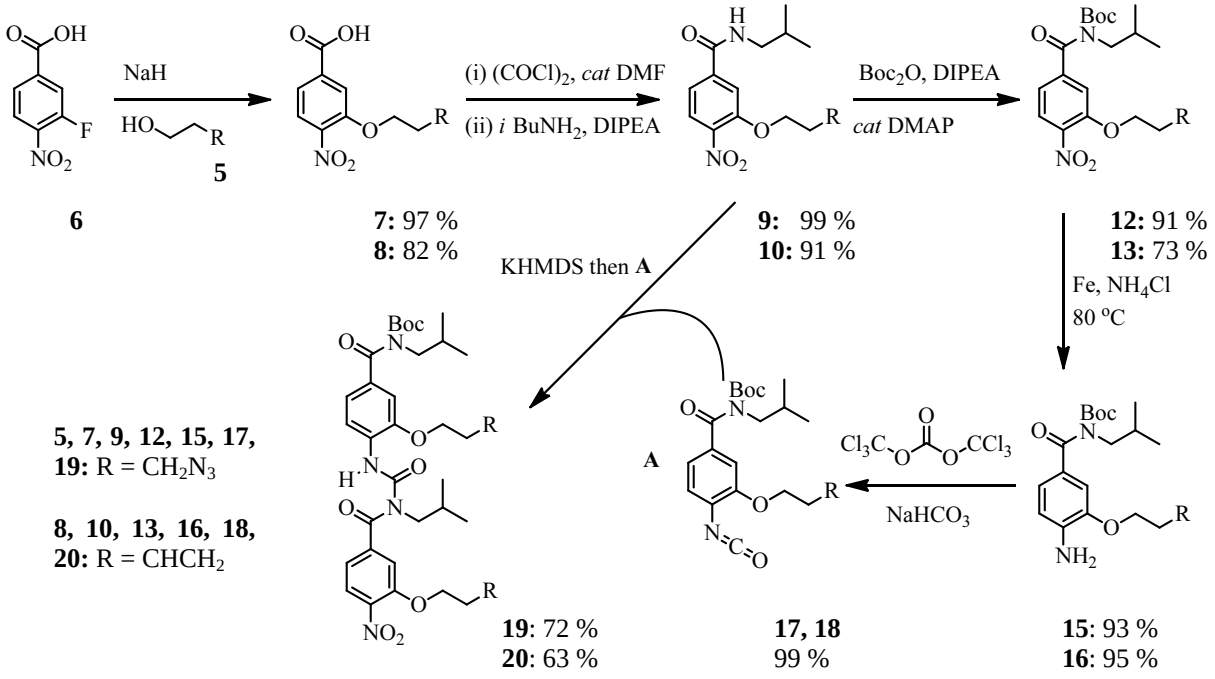
Overlaying a thermodynamically-accessible conformation of the proposed mimetic **1** (calculated by a semi-empirical optimization) on the crystal structure of Gab2a indicates that the RMSD of the C_α and C_β atoms of both structures is 0.37 Å (Figure 3-3). The scaffold also retains significant flexibility, implying that it may bind *via* an "induced fit" mechanism. With this scaffold selected, functionalization was possible along several points: the identity of the polar sidechains, the identity of the aliphatic sidechains, and the number of benzoylurea linkages employed. In the first instance, the simplest benzoylurea capable of

incorporating the key Arg and Lys sidechains was selected. This consists of a single urea linkage, two substituted phenyl rings, and two amides. Molecule **1** thus features Arg and Lys sidechains to match the critical residues in the Gab2a peptide. Isobutyl groups are incorporated to provide hydrophobic contact with the protein. These are intended to mimic the essential proline residues in the peptide, though they may not be ideal for this purpose as they are aliphatic groups instead of *N*-substituted rings. Isobutyl groups, however, are much easier to incorporate synthetically, and may be adequate to allow binding. Directionally, the molecules is synthesized with Arg on the "top" ring to allow the isobutyl groups to best overlay with prolines in the peptide (Figure 3-3). Compound **2**, a simplified model of **1** with two Lys sidechains, was selected as an initial target for synthesis. Additionally two control molecules, diol **3** and aliphatic **4**, investigate the importance of hydrogen bonding to binding affinity for Grb2 SH3C.

3.3. Synthesis of 1st generation mimics

Previous work on the benzoylurea scaffold has only incorporated hydrophobic and anionic sidechains.[16–18] This project requires the more challenging cationic Lys and Arg sidechains. Due to the difficulty of purifying and handling primary amines and guanidines, the synthesis of the target molecules was designed such that the these groups remained protected until a late stage. An azide was chosen as the protected form for the Lys sidechains of **2** as its presence could be easily detected *via* its distinctive IR stretch and it was expected to be easily reduced *via* hydrogenation. The control molecules were derived from an alkene, which could be transformed *via* hydrogenation to the aliphatic chains of **4**, or *via* ozonolysis followed by reduction to yield diol **3**. Furthermore, the alkene functional group is amenable to a variety of functional group conversions, allowing facile access to a host of other compounds.

A robust, scalable, and functional group-tolerant approach to benzoylurea mimetics from an isocyanate and a secondary amide has previously been developed by the Hamilton group.[16,17] Key to this synthesis is the protection of all acidic hydrogens except for that of the secondary amide. The synthesis of both precursors began with 3-fluoro-4-nitrobenzoic acid **6** (Scheme 1). The azidoalcohol **5** was synthesized by displacement of bromide from 3-bromo-propanol with sodium azide. Nucleophilic aromatic substitution then set the first of the sidechains (**7**, **8**). After the formation of an acid chloride with oxalyl chloride, amide bond formation with isobutyl amine provided the lower half of the molecule (**9**, **10**).

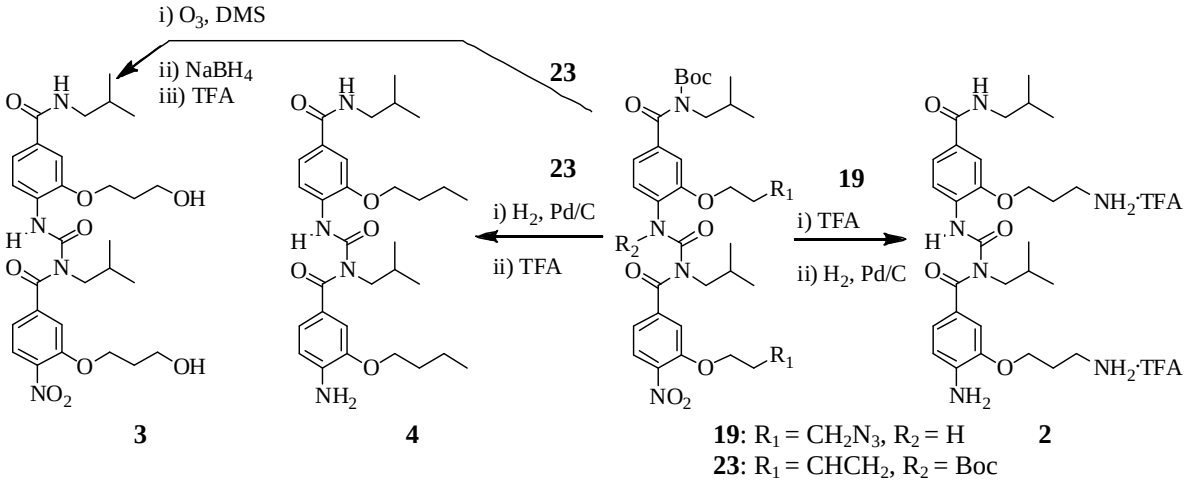


Scheme 3-1: Synthesis of the benzoylurea scaffold.

The route to the isocyanate proceeded with the protection of the amide as the *tert*-butyl carbamate (**12**, **13**), followed by reduction of the nitro group. This reduction proved challenging in the presence of azide **12**, as the majority of published methods for reducing aromatic nitro groups cause concomitant reduction of azides. Ultimately, iron and ammonium chloride in refluxing ethanol and water was found to selectively reduce the nitro group. Once this reduction was complete, the azide began to reduce under the same

conditions, but if the reaction was stopped promptly aniline **15** was formed in excellent yield. The same reaction was applied to the alkene substrate to form aniline **16**. Isocyanate formation then proceeded in quantitative yield using triphosgene and sodium bicarbonate in biphasic DCM/water (**17**, **18**). With both fragments now in hand, the amide was deprotonated using potassium hexamethyldisilazide in THF at -78 ºC before dropwise addition of the isocyanate.

From di-azide benzoylurea **19**, the di-lysine mimic **2** was formed by sequential Boc deprotection using TFA and hydrogenation with palladium on activated carbon (Scheme 3-2). At this stage, purification could only be effected using semi-preparative HPLC. Even using this method, the product proved difficult to separate from an unidentified contaminant which eluted very closely with the product. Eventually, separation was achieved by injecting relatively dilute samples (~25 $mg.mL^{-1}$), but this required several injections in order to purify a sufficient quantity of material **2** for biological testing. A final counterion exchange using 0.1 M HCl could also produce the product as the HCl salt **22**.

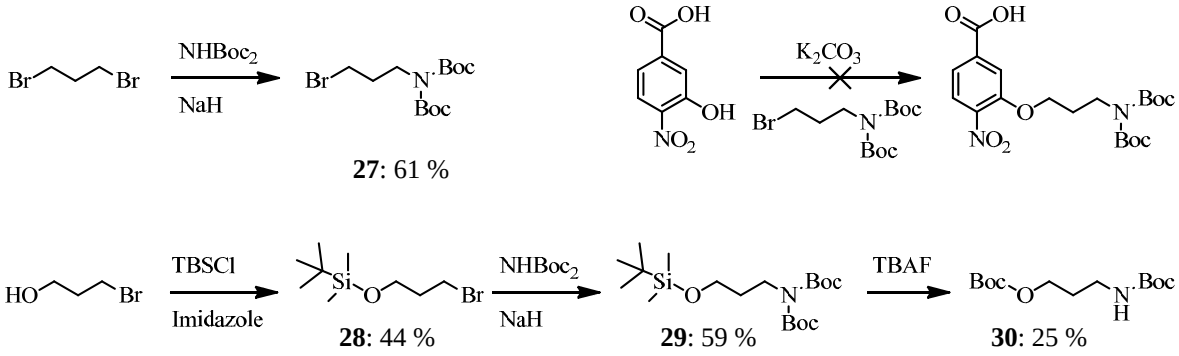


Scheme 3-2: Synthesis of mimics **2**-**4** from the coupled benzoylureas **19** and **23**.

The control molecules were synthesized in a divergent strategy from diene **20** (Scheme 3-2). The benzoylurea nitrogen was first protected as the *tert*-butyl carbamate **23**. For the aliphatic compound **4**, the olefins were then reduced *via* hydrogenation to alkane **24**,

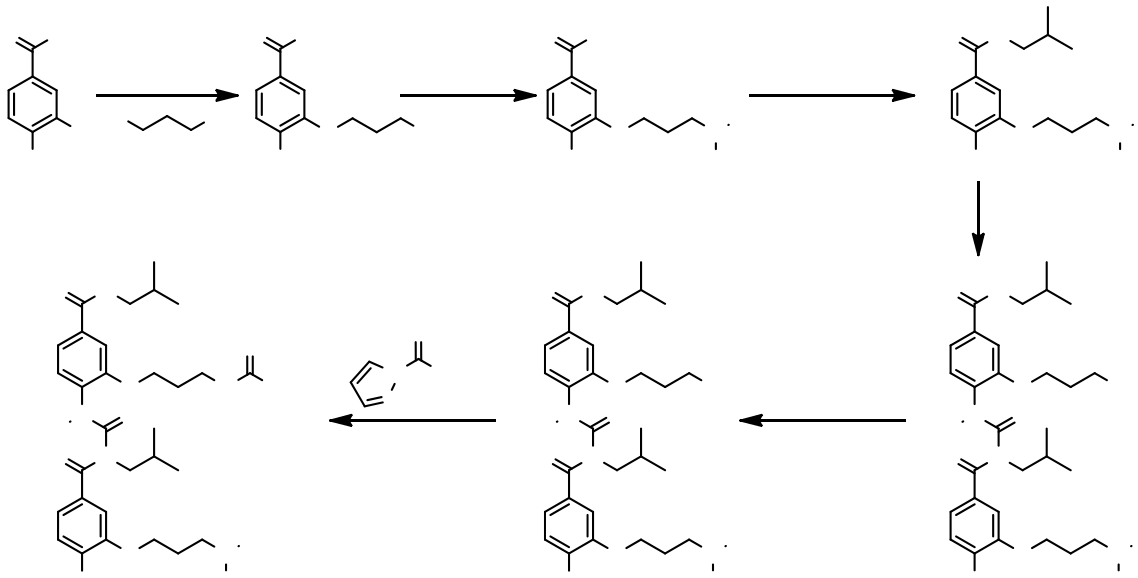
followed by removal of the Boc groups with TFA. To form diol **3**, the olefins were cleaved *via* ozonolysis and dimethylsulfide workup to yield di-aldehyde **25**. Reduction with sodium borohydride yielded the diol **26**, which was subsequently deprotected with TFA to the final product **3**.

The more elaborate Arg/Lys molecule required two orthogonal protecting groups to form the different sidechains. Moreover, due to the harshly basic conditions utilized in the coupling step, all acidic protons had to be protected and the desired amine and guanidine groups masked. An azide was chosen for the isocyanate half of the molecule, and Boc protection was pursued for the amide portion.



Scheme 3-3: Attempted syntheses of *N*,*N*-diBoc protected sidechains

Standard Boc protection conditions would allow for the introduction of only a single Boc group, so it was decided to try to directly incorporate an *N*,*N*-di-Boc protected amine (Scheme 3-3). 1,3-di-Bromopropane was converted to 1-bromo-3-di-Boc-amine **27** in one step with the goal of a substitution reaction with 3-hydroxy-4-nitrobenzoic acid, but no reaction was observed. Efforts were then made to synthesize di-Boc-amine propanol for subsequent S_NAr reaction. 3-Bromo-propanol was TBS protected to give **28**, and the bromide displaced with the anion of *N*,*N*-di-Boc-amine to form **29**. This was followed by TBAF deprotection of the alcohol intended to yield the free propanol, but instead Boc transfer occurred under the reaction conditions to form the mono-protected amine and Boc protected alcohol **30**.



the azide group to primary amine **35**. This was purified by filtration through Celite® under neutral conditions to prevent protonation of the primary amine. The free amine of **35** was sufficiently nucleophilic to react with a protected guanidine group (*N*,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine) to furnish the penta-Boc protected **36**. Deprotection with TFA yielded the final molecule **1**, which was purified by reverse-phase HPLC.

3.4. Biological evaluation of 1st generation mimics

3.4.1. Surface Plasmon resonance

A surface plasmon resonance assay was the first step in probing the binding of the mimics. Unlabelled Grb2 SH3C was covalently bound to the dextran surface of a CM5 chip by amine coupling chemistry. The small molecule ligands were then dissolved in running buffer and screened for binding at 50 μM and 100 μM concentrations.

Table 3-1: SPR binding response (given in arbitrary Response Units) for mimics **1**-**4** screened at 50 μM and 100 μM concentrations. Molecule **1**, featuring Arg and Lys sidechains, shows a much greater response than any of the other mimics.

	1	**2**	**3**	**4**
50 μM	111.8	14.2	3.5	11.7
100 μM	194.8	23.8	5.4	13.2

Molecule **1**, which showed the most promising response by SPR, was titrated in a two-fold dilution series at concentrations from 3.125 μM to 100 μM. Above this concentration, undesirable superstoichiometric binding began to occur, indicative of possible aggregation or non-specific interaction with the protein or dextran surface. The resulting maximal responses were fitted to a one-to-one binding model:

$$Response(RU) = B_{\max} \frac{[I]}{K_d + [I]}$$

where [*I*] is the concentration of the inhibitor, B_{max} is the maximal binding response, and K_d is the inhibitor concentration yielding half-maximal binding. This model calculated

K_d = 240 ± 20 μM, or approximately one quarter of the 58 μM binding affinity of the native Gab2a PPII peptide (Figure 3-4). Given that the small molecule weighs ca. 600 Da vs. ca. 1250 Da for the peptide, this represents a ligand efficiency of 50 % that of the native binder.

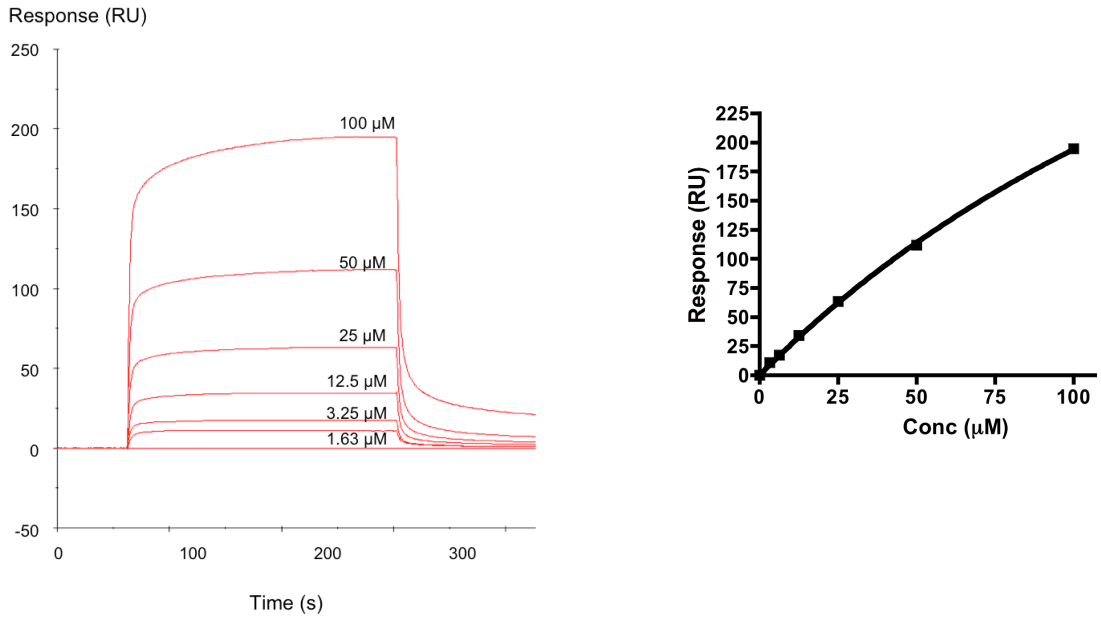


Figure 3-4: Left: sensograms of the binding of mimic **1** to Grb2 SH3C. Right: fit of binding response vs. mimic concentration to a one-to-one binding model. This model calculates an affinity of 240 ± 20 μM.

3.4.2. Competition binding assay

Results from SPR were validated using the competition binding assay described in Chapter 2. GST–Grb2 SH3C was incubated with each mimic (2.0 mM) before contact with streptavidin beads coupled to biotinylated Gab2b peptide. Solubility of control molecule **4** was slight at this concentration in aqueous buffer with 5 % DMSO. This sample was included in the analysis although the true concentration of mimic in solution could not be estimated accurately. A threshold of 20 % minimum inhibition was applied to eliminate ambiguous-, weak-, or non-binders. Using this criterion, only the Arg/Lys mimic **1** was found to bind significantly (Figure 3-5). This positive result confirms that the mimic both binds to Grb2 SH3C, and can competitively inhibit peptide binding.

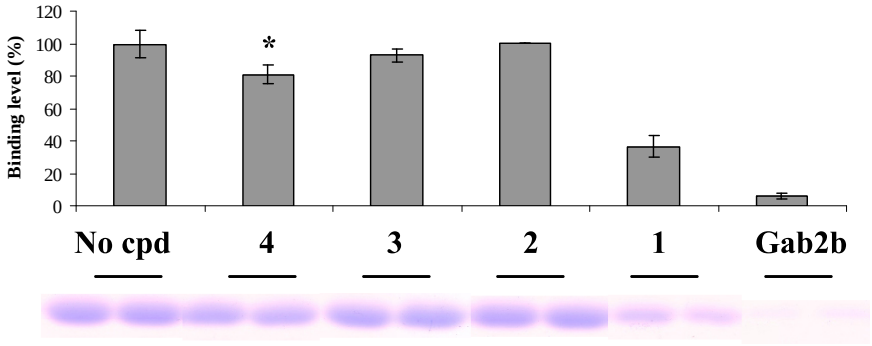


Figure 3-5: Normalized intensities (top) and Coomassie blue-dyed bands (bottom) for the competition assay using the 1st generation mimics. Samples are scaled relative to the negative control (left column) which does not include any compound. Mimic **4** (starred) shows a small inhibitory effect, but below the 20 % threshold established to separate binders from non-binders. This mimic was minimally soluble in the assay conditions so its concentration cannot be estimated accurately. Only the Arg/Lys mimic **1** competitively inhibits the interaction. The positive control (right column) is a ten-fold excess of unbound Gab2b peptide.

3.5. Design and synthesis of 2nd generation mimics

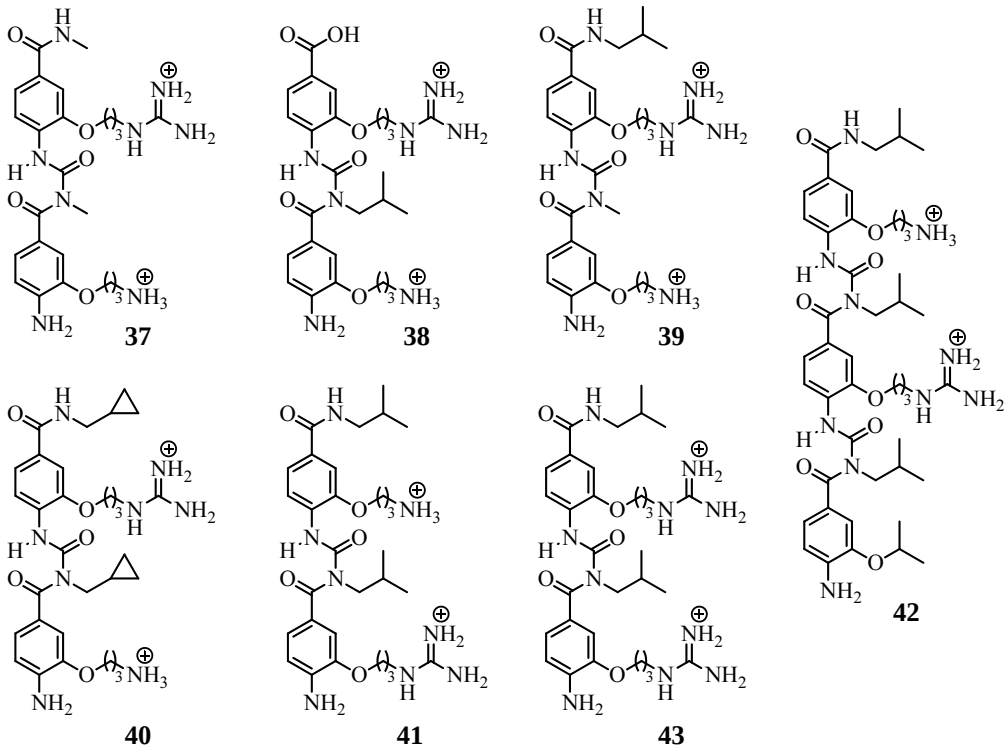
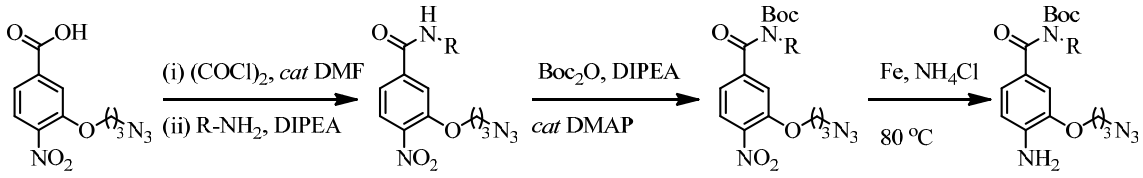


Figure 3-6: Targeted 2nd generation peptidomimetic library.

With the success of mimic **1** by SPR and the competition assay, a focused 2nd generation library was designed to further probe the binding site specificity of the protein (Figure 3-6). The 1st generation molecules revealed the Arg and Lys sidechains to be essential to binding, so the new molecules all incorporated that functionality. Instead, the aliphatic sidechains

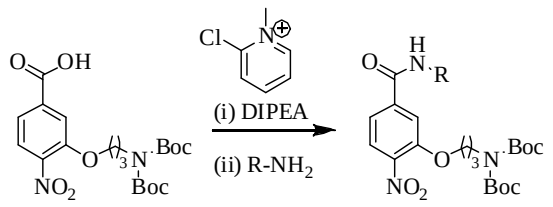
were systematically varied: **37** replaces both isobutyl groups with smaller methylamides. Mimic **38** eliminates the first of the isobutyl sidechains, replacing it instead with the free acid, while mimic **39** retains the first isobutylamide and replaces the second with methylamide. In **40**, both sidechains are cyclic cyclopropylmethyl groups. Molecule **41** reverses the pseudo-*N*- and *C*- termi of **1**, to investigate the directionaly of the scaffold. The extended mimic **42** adds a second urea linkage and two additional aliphatic sidechains to ideally pick up additional hydrophobic contacts. Molecule **43** contains two Arg sidechains with isobutyl groups. Additionally, a molecule featuring two isopropyl sidechains was planned to investigate the role of α-substitution on binding.



6	R = *i*Pro	**45**: 54 %	**48**: 26%	**52**: 52 %
	R = Me	**46**: 96 %	**49**: 73 %	**53**: 92 %
	R =cProMe	**47**: 81 %	**50**: 86 %	**54**: 91 %

Scheme 3-5: Synthesis of "top" fragment with different amides.

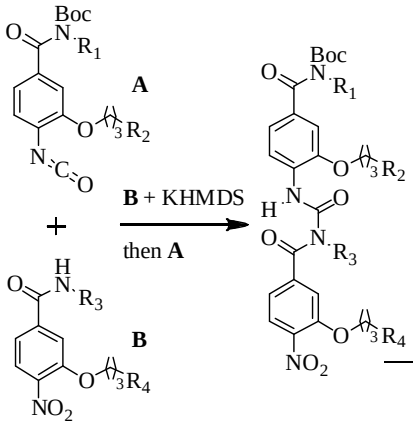
Synthesis of these molecules proceeded through a nearly identical route as for molecule **1**. The azidoalcohol **5** was carefully prepared on a larger scale (2.7 g) and handled behind a blast shield until converted to the more stable **7** by S_NAr reaction. This acid was either protected as the *tert*-butyl ester **44** (for the eventual acid of **38**) or converted to an amide by synthesis of the acid chloride followed by reaction with the appropriate amine. The additional steps of *N*-Boc protection, nitro group reduction, and isocyanate formation proceeded in reasonable to excellent yields (Scheme 3-5).



	Amide
R = *i*Pro	**59**: 37 %
R = Me	**60**: 61 %
R = cProMe	**61**: 69 %

Scheme 3-6: Synthesis of the "bottom" fragment with different amides.

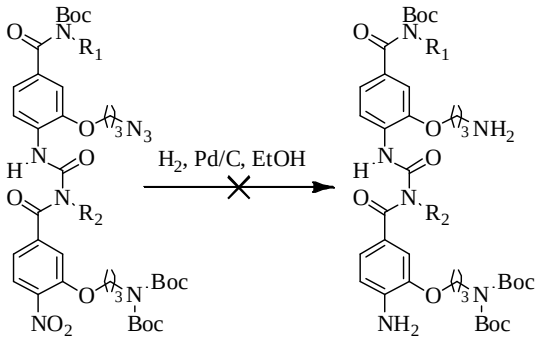
The bottom fragments were again prepared through the common *N*,*N*-di-Boc-amine intermediate **32**, from which the desired amides were synthesized using Mukaiyama amide coupling conditions (Scheme 3-6). The "reverse" molecule **41** was synthesized by analogous means, but this time with the identity of the sidechains reversed. Fragment **33**, used as the bottom half of **1**, was instead *N*-Boc protected to **62**. The aromatic nitro group was reduced to aniline **63** by hydrogenation and the isocyanate **64** formed under biphasic conditions with triphosgene and $NaHCO_3$.



		Isocyanate	Amide	Benzoylurea
$R_2 = N_3$ $R_4 = NBoc_2$	$R_1 = R_3 =$ *i*Pro	**56**	**59**	**65**: 48 %
	$R_1 = R_3 =$ Me	**57**	**60**	**66**: 90 %
	$R_1 = R_3 =$ cProMe	**58**	**61**	**67**: 78 %
	$R_1 =$ O*t*Bu $R_3 =$ *i*Bu	**55**	**33**	**68**: 91 %
	$R_1 =$ *i*Bu $R_3 =$ Me	**17**	**60**	**69**: 64 %
$R_2 = NBoc_2$ $R_4 = N_3$	$R_1 = R_3 =$ *i*Bu	**64**	**9**	**70**: 81 %

Scheme 3-7: Coupling of benzoylureas.

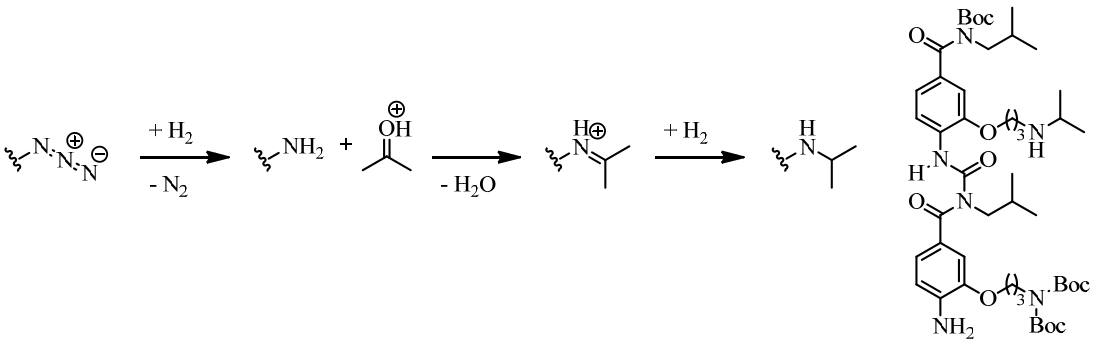
For all of the molecules, coupling was conducted by deprotonating the amide with KHMDS in THF at -78 ºC and adding the isocyanate dropwise (Scheme 3-7). If excess base was used, yields were adversely affected by the formation of the elimination product in which di-Boc-amine leaves to yield an alkene. Stoichiometric base reduced this to a trace impurity that was easily separated *via* column chromatography. While yields for these reactions were in general very good, the synthesis of the molecule including isopropyl sidechains was less successful. This was rationalized as an effect of the greater steric hindrance of the α-substituted isopropyl sidechain relative to the other molecules (as evidence by the A value for isopropyl of 2.2 vs. 1.7 for methyl).[21] While coupling was successful to form the di-isopropyl molecule **65**, when this product was lost in subsequent reactions it was not resynthesized.



	Benzoylurea	Product
$R_1 = R_2 =$ *i*Pro	**65**	Incomplete
$R_1 = R_2 =$ Me	**66**	Incomplete
$R_1 = R_2 =$ cProMe	**67**	Incomplete
$R_1 =$ O*t*Bu $R_2 =$ *i*Bu	**68**	Incomplete
$R_1 =$ *i*Bu $R_2 =$ Me	**69**	Incomplete
$R_1 = R_2 =$ *i*Bu	**34**	**35**: *quant*

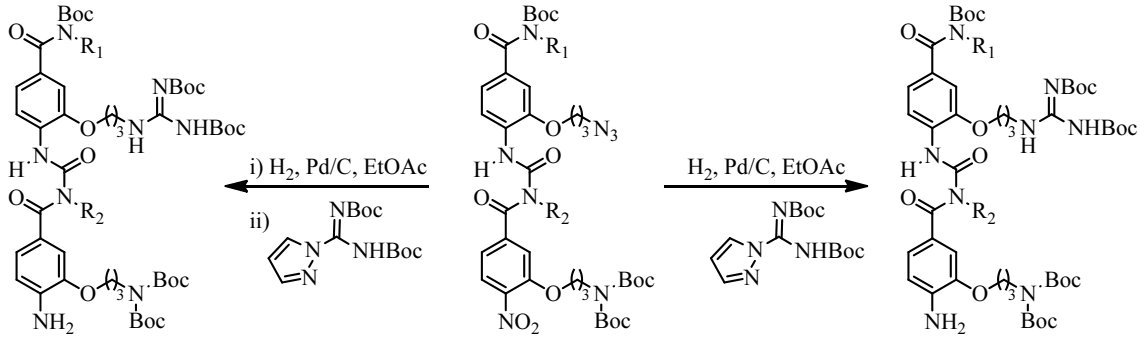
Figure 3-7: Attempted hydrogenation of the azide of the coupled benzoylurea to the primary amine. Although these conditions were successful for the 1st generation molecule **34**, they failed for all of the 2nd generation mimics.

At this stage, molecules **65-70** were subjected to hydrogenation under the conditions that were successful for **35**: Pd/C in ethanol overnight under a balloon of hydrogen. This reaction, which had previously proceeded in quantitative yield, failed to produce pure product for these new substrates (Figure 3-7). Purification was not possible at this stage, so conditions were investigated for driving the reaction to completion. Crude ^{1}H NMR showed mixtures of products, while crude IR indicated that the azide had been reduced to the amine. This difference in behaviour caused by changes in the aliphatic portions of the molecules could not be explained. Changing the reaction solvent to ethyl acetate resulted in complete reaction of substrates **67** and **70**, but not of the other closely-related molecules.



Scheme 3-8: Formation of the secondary amine **73** by condensation with acetone. **73**: 96 %

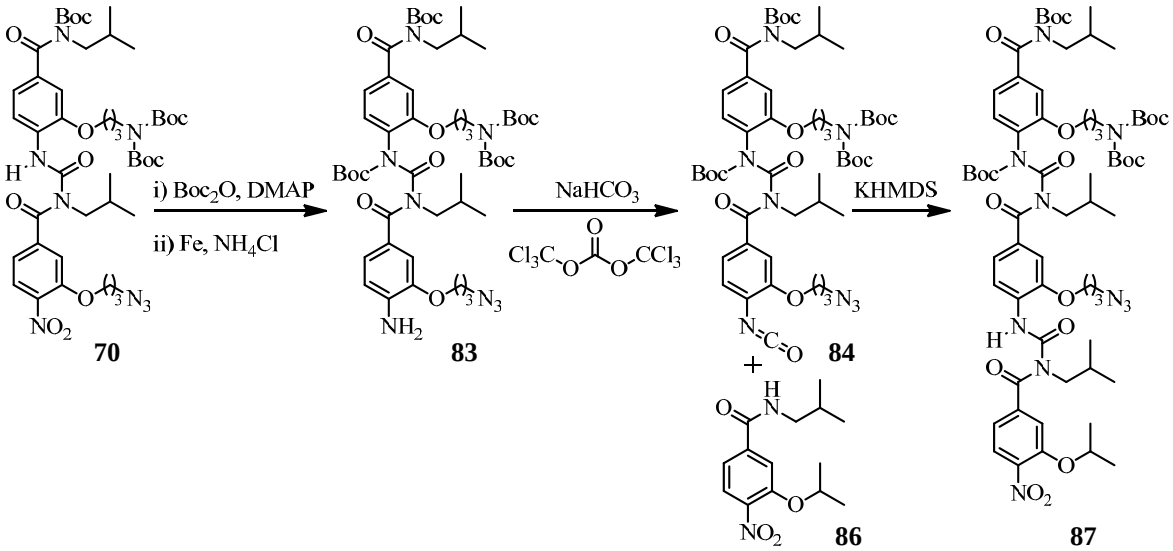
Changing catalyst to $Pd(OH)_2$ or PtO_2, increasing the catalyst loading to stoichiometric Pd/C, adding acetic acid, and heating all failed to produce pure product. A solvent screen (of THF, toluene, AcOH, acetone, DCM, DMF, and EtOAc) was similarly inconclusive, though when the reaction was conducted in acetone the secondary amine **73** was formed, presumably in a stepwise process of hydrogenation to the amine followed by reductive amination with solvent (Scheme 3-8). The reaction was conducted in wet acetone and Pd/C is frequently acidic, so the reaction may have proceeded through the iminium ion. Despite the three steps required, this reaction formed pure **73** by NMR. This inspired consideration of a mechanism by which the primary amine product of the hydrogenation somehow poisons the catalyst or reacts with other substrate molecules. If the amine could be removed from the reaction mixture as in the case of the reductive amination, the reaction might reach completion.



Scheme 3-9: Formation of protected Arg sidechain in two steps (left) or one step (right).

In a modification of the original procedure the protected guanidinium reagent was added to the hydrogenation mixture and the reaction stirred for two to four days (Scheme 3-9). This led to a single product peak by TLC and excellent yields in one step for the formation of the protected Arg sidechains from the azides. When this modified procedure was tested on the original substrate **34**, it resulted in a significantly greater yield and avoided any

problems with the possible protonation of **35**. Finally, products **73-78** were deprotected in TFA and purified *via* reverse-phase HPLC to yield final molecules **37-41** and **79**.



Scheme 3-10: Synthesis of the trimeric benzoylurea **42**.

Di-Arg molecule **43** was synthesized from di-Lys precursor **80** by treatment with the guanidiunium reagent followed by global deprotection. The extended mimic **42** was synthesized from the coupled "reverse" molecule **70** (Scheme 3-10). The benzylurea nitrogen was Boc protected and the aromatic nitro group reduced using iron and ammonium chloride to yield **83**. The isocyanate **84** was formed, before coupling with the aliphatic benzamide **86** to yield double benzoylurea **87**, incorporating six sidechains. The Arg sidechain was installed in the modified one step procedure from the azide, resulting in the sextuple-Boc protected **88**. Global deprotection with TFA yielded the extended mimic **42**, which was purified by reverse-phase HPLC.

3.6. Biological evaluation of 2nd generation mimics

3.6.1. Surface plasmon resonance

The analysis of the 2nd generation mimics adapted the procedure used for the earlier molecules. The only modification was the use of two separate Grb2 SH3C analysis surfaces, one containing the protein construct used for the prior work, and the second containing the

^{15}N-labeled version of the same protein. The constructs were covalently attached to the dextran surfaces to a level of 1100 RU. Those molecules which were water-soluble (all but **42** and **43** of the 2nd generation compounds) were screened in running buffer without DMSO, while in a second run, 5 % DMSO was added (and the peptide controls repeated).

Table 3-2: SPR binding response for 2nd generation mimics screened at 50 μM and 100 μM concentrations. All except mimic **38** show significant binding, but with varying levels of response. The total binding levels are higher for the unlabelled surface, but the ratio of the response to the labeled vs. unlabeled surfaces is not consistent.

Molecule	**Unlabelled Grb2 SH3C**		**^{15}N Grb2 SH3C**	
	50 μM	**100 μM**	**50 μM**	**100 μM**
43	82.3	136.6	46.7	76.1
37	55.0	83.1	25.5	39.2
38	5.0	11.2	4.1	8.5
39	52.2	95.6	45.0	73.7
79	25.9	43.6	20.6	34.5
40	48.6	88.1	36.0	59.6
41	36.2	62.3	24.3	41.4

As expected from the targeted nature of these molecules, nearly all bound to Grb2 SH3C at the 50 μM and 100 μM concentrations (Table 3-2). Although nearly equal amounts of protein were bound to the two surfaces, substantial differences, often exceeding 50 %, were observed for the binding to the labeled and unlabeled constructs. Moreover, while binding was always greater to the unlabelled protein, the ratio of the response to the two flowcells was not constant. For instance, for **37** the binding to the two cells differed by a factor of two, while for **39** the ratio was just 1.2:1. This cannot be explained by differences in the amount of protein bound or by differences in purity, since in either case the relative responses should be consistent. The extended mimic **42** displayed poor peak shape, and so was excluded from SPR analysis.

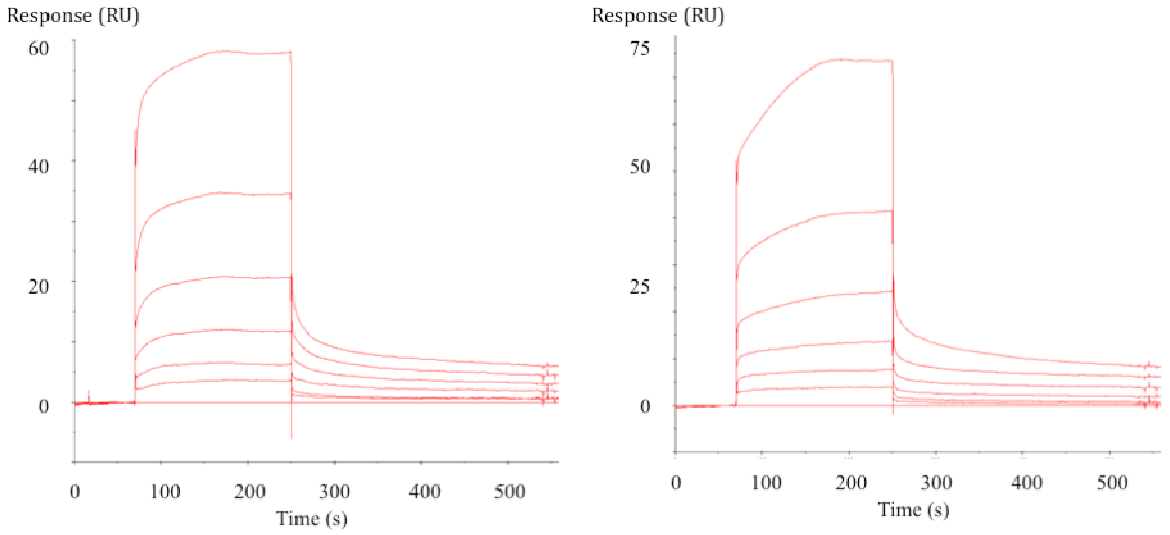


Figure 3-8: Sensograms of mimics **79** (left) and **40** (right) binding to the surface containing ^{15}N-labeled Grb2 SH3C. Mimic concentrations are a two-fold dilution series from 200 μM (highest response) to 6.25 μM (lowest non-zero response). For **40**, the 200 μM response is more than twice that of the 100 μM response, indicating that non-stoichiometric binding is occurring at the highest concentration. As a result, the binding analysis (Table 3-3) excludes this point.

To quantify the binding of the mimics, serial titrations from 6.25 μM to 200 μM were collected for each molecule (Figure 3-8). Fits were performed separately for the ^{15}N and unlabelled flowcells to the one-to-one model described previously (Table 3-3). By SPR, the binding affinity of the control Gab2b peptide was calculated to be slightly higher than determined by ITC (15 μM and 11 μM, respectively, for the unlabelled and ^{15}N constructs vs. 3 μM by ITC). More worryingly, the 1st generation molecule **1** was calculated to bind with nearly three-fold greater affinity than before, with a K_d of approximately 90 μM.

Table 3-3: Affinities of the 2nd generation mimics calculated from SPR titrations. Values are listed with their statistical errors from non-linear least-squares fits to a one-to-one binding model. NC indicates that the fit to the data did not converge to a finite value. For the starred mimics **39** and **40**, the sensograms displayed superstoichiometric binding at the highest concentration of 200 μM, so the fits exclude that point. Molecule **1** (dagger) displayed superstoichiometric binding at 100 μM, so the fit is for a two-fold dilution from 60 μM to 3.75 μM.

Molecule	Unlabelled Grb2 SH3C K_d (μM)	^{15}N Grb2 SH3C K_d (μM)
1	90 ± 20†	83 ± 6†
43	NC	NC
37	150 ± 40	200 ± 30
38	NC	NC
39	320 ± 70*	160 ± 10*
79	500 ± 100	320 ± 40
40	250 ± 60*	150 ± 20*
41	800 ± 300	450 ± 80

Analysis of the ligand structures reveals plausible SAR. Removal of the "top" amide abolished binding, with molecule **38** displaying unquantifiably-weak binding. The second hydrophobic group was less important, with the methyl truncate **39** binding just two-fold weaker than the original molecule **1**. Replacing both groups with methylamides in **37** did not cause additional loss of binding affinity. The cyclized sidechains of **40** did not improve binding relative to the mono-truncate **39**, and remained less active than **1**. Interestingly, the reverse molecule **41** bound very weakly, despite having the same sidechains as **1**. Taken together, this SAR suggests that binding of the benzoylurea is directional, with the Arg and Lys sidechains occupying defined positions in the binding site, perhaps fixed by two hydrogen bonds from the Arg sidechain to E16 of Grb2 SH3C. This places the hydrophobic groups in defined positions. The "top" amide provides key hydrophobic contacts to a hydrophobic pocket defined by F9 and Y51. The reverse molecule **41** fails to occupy this site, and as a result binds weakly.

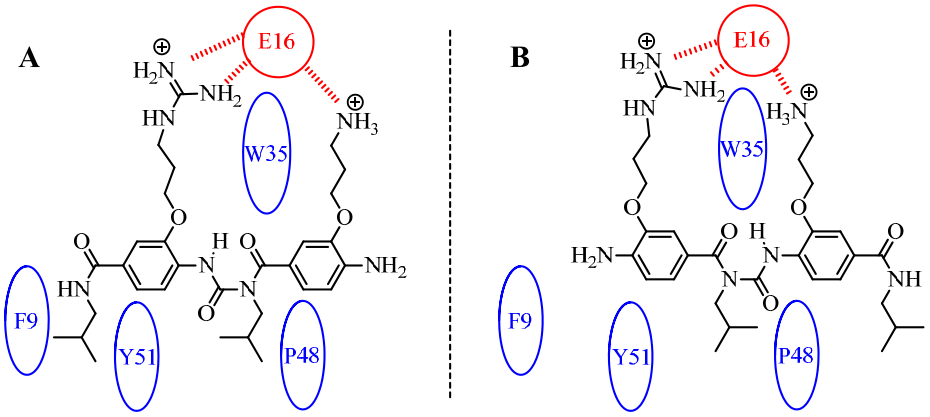


Figure 3-9: Proposed binding modes for (a) the Arg/Lys mimic **1** and (b) the "reverse" mimic **41**. Blue indicates a hydrophobic interaction, and red indicates a hydrogen bond. In both cases, the location of the Arg sidechain is fixed by two hydrogen bonds to E16. This positions the hydrophobic groups into defined pockets. The original mimic **1** fills an aromatic pocket containing F9 and Y51, while the reverse mimic is unable to occupy this position. This results in substantially weaker binding of **41** relative to **1**.

While this appears logical, inconsistencies within the SPR data require confirmation in other assays. In particular, the discrepancies in binding between the two constructs, the irreproducibility of the binding constant of **1** relative to past experiments, and the poor peak shape for some of the mimics is of concern.

3.6.2. Competition binding assay

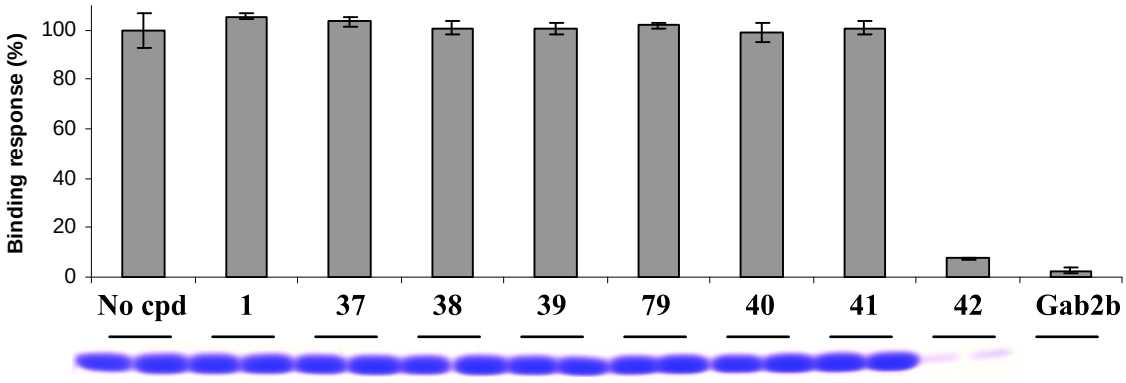


Figure 3-10: Normalized intensities (top) and Coomassie blue-dyed bands (bottom) for the competition assay using the 2nd generation mimics. Significant inhibition is found with mimic **42**. In this assay, however, the previously-successful mimic **1** fails to show any inhibition. The positive control Gab2b peptide is in 100-fold excess.

The orthogonal competition assay was conducted with each of the molecules to further assess their binding. Consistent with the testing of the 1st generation molecules, the mimics were tested at 2.0 mM concentration. This time, although both the positive and negative

controls functioned as expected, the original hit **1** failed to bind. Indeed, only the extended mimic **42** showed substantial knockdown of Gab2b binding (Figure 3-10).

To quantify the competitive component of this binding, the competition assay was repeated with increasing concentrations of compound **42**. Using this assay, no clear dose-response behaviour was observed, indicating that the binding does not result from simple competitive binding to the active site (Figure 3-11). The increased protein levels detected at with 1.25 mM **42** must be erroneous (despite the close agreement of the technical duplicates), but the lack of additional response at 2.5 mM compound implies that properties such as aggregation or nonspecific binding may be responsible for the activity seen.

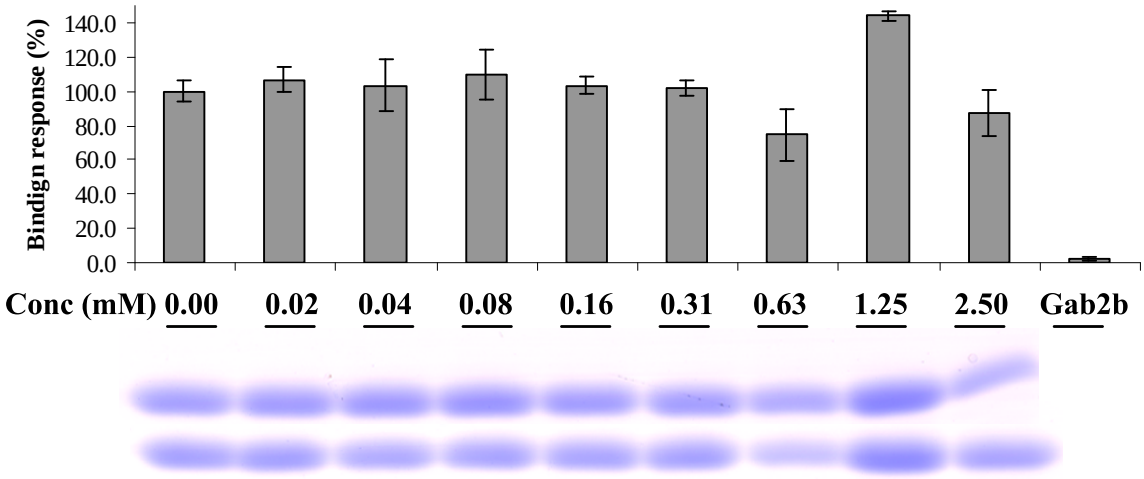


Figure 3-11: Normalized intensities (top) and Coomassie blue-dyed bands (bottom) for the competition assay using increasing concentrations of mimic **42**. While an inhibitory effect begins at a concentration of 0.63 mM, it does not produce clean behavior at higher concentrations. The positive control Gab2b peptide is in 100-fold excess.

3.6.3. NMR: ^{15}N HSQC shifts

Given the inconsistencies of the SPR and competition binding assays, protein NMR experiments were conducted. Uniformly ^{15}N-labelled Grb2 SH3C was expressed and purified (by Philip Simister) for protein-observed ^{15}N HSQC shift experiments. In this experiment, the ^{1}H chemical shifts of backbone (and select sidechain) amide protons are correlated to the ^{15}N chemical shifts of the amides. These shifts are very sensitive to conformation and local chemical environment (including hydrogen bonding to the residue

or adjacent carbonyl), so changes in the protein structure are readily detected.[22] If a ligand binds to a protein, upon titration it will cause a chemical shift change for both the ^{1}H and ^{15}N nuclei, with the magnitude of the shift depending on the degree of perturbation. This experiment thus gives information both on the location of the binding site and on affinity, if titrations are performed over a large enough range of ligand concentrations.

Importantly, NMR is limited in the range of affinities it can detect, as exchange must be fast on the NMR timescale or else peaks for the "bound" and "unbound" forms resolve separately. Equilibrium kinetic analysis shows that K_d is equal to the "on" rate of a reaction divided by the "off" rate. Assuming that protein-ligand complex formation is diffusion-limited ($k_{on} \approx 10^9\ M^{-1}s^{-1}$),[23] $k_{off} \approx 10^9\ K_d$. If the maximum chemical shift change observed is 0.1 ppm for ^{1}H on a 700 MHz magnet (70 Hz), this corresponds to a detectable $k_{off} \approx 70\ s^{-1}$, so if a binding event is "fast" (10x) compared to this, the affinity must be no greater than ~ 1 μM.

Due to the limited sensitivity of NMR, high concentrations of both protein and ligand were required. Using a 700 MHz magnet and cryoprobe, 200 μM ^{15}N Grb2 SH3C gave a high-quality HSQC spectrum in approximately 20 minutes. The signal-to-noise ratio for NMR varies with the square of sample concentration, so lower concentrations required substantially more instrument time. The ^{15}N HSQC shifts of Grb2 SH3C have been reported,[24] and the initial spectrum of our construct matched these resonances closely, though with the addition of lower-level contaminant peaks (Figure 3-12). Although the identity of these peaks cannot be conclusively determined, circumstantial evidence points to a partially-cleaved and unfolded version of Grb2 as the contaminant contains two Trp, and has many resonances in the unstructured region of the NMR spectrum.

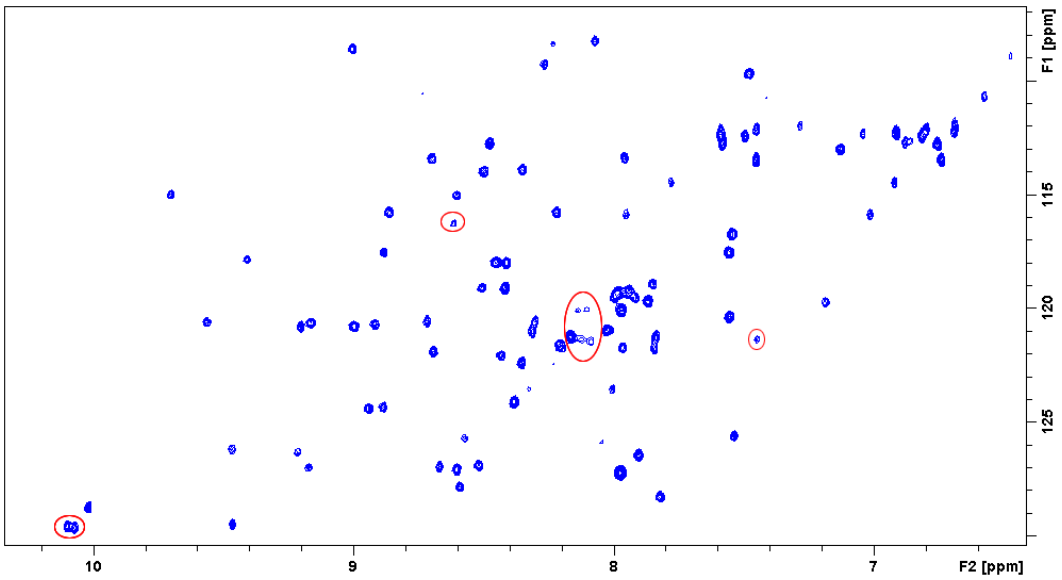


Figure 3-12: ^{15}N HSQC spectrum of Grb2 SH3C at pH 7.2. The resonances are well-resolved, and closely match those reported in the literature. Selected residues from the cleaved species are circled in red.

Since NMR resonances can be sensitive to pH, initial control spectra were recorded at pH 7.2, 7.4, 7.6, and 7.8 in phosphate buffer. While for many proteins amide resonances broaden and disappear at higher pH, well-resolved spectra were recorded even at the highest pH. Between pH 7.2 and 7.4, several residues shifted in both ^{1}H and ^{15}N resonances, but very little additional change occurred at higher pH. The addition of 5 % DMSO-d6 at pH 7.4 resulted in no shifts, confirming that the protein is stable in the presence of some organic solvent.

To validate the assay, unlabelled Gab2b peptide was added in six-fold molar excess. This resulted in numerous large resonance shifts (> 0.1 ppm) and smaller shifts to many additional residues (Figure 3-13). The perturbed residues map closely to those which contact the peptide in the crystal structure,[25] with the exception of the less-structured *N*-terminus of the peptide, which perturbs the protein less than the cutoff amount.

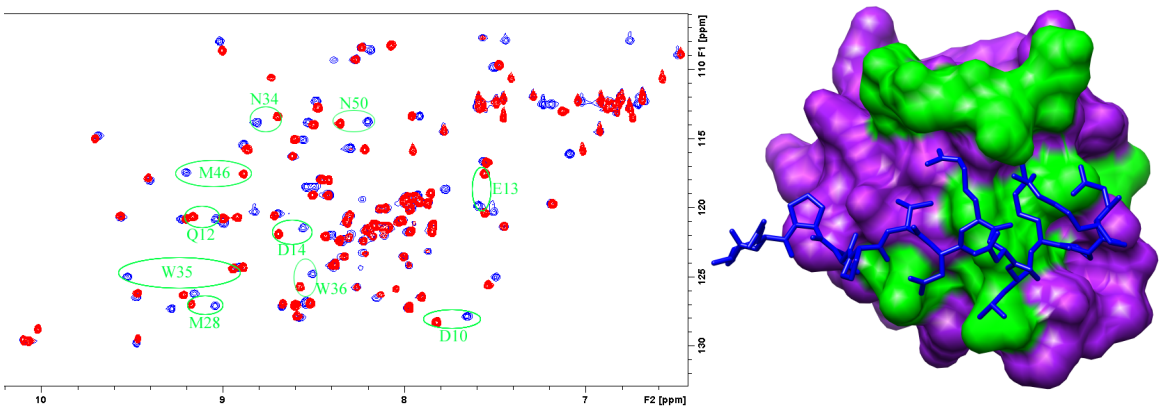


Figure 3-13: Left: ^{15}N HSQC spectrum of Grb2 SH3C (red), overlaid with the spectrum upon addition of 6 eq of the Gab2b peptide (blue). The peptide perturbs a number of residues (shifts of > 0.1 ppm highlighted and labeled in green). Right: perturbed residues (green) mapped to the crystal structure of Grb2 SH3C binding the Gab2b 3_{10} peptide (2VWF; blue). The residues perturbed in the NMR experiment match closely to the residues contacted in the crystal structure.

With the NMR assay validated, the experiment was repeated with the 1st generation mimic **1**. In a first attempt, one molar equivalent was added to buffered protein. A white precipitate formed immediately, exactly at the tip of the pipette that was adding the small molecule. As the compound was soluble in water at a fifty-fold higher concentration than was used in the NMR assay, and as this behaviour had not been seen previously, it was hypothesized that it was an effect of high protein concentration. Grb2 SH3C is relatively insoluble at low pH, so a localized pH change combined with high protein concentration could explain the behaviour seen. A small amount of an aqueous solution of **1** was added to pH paper, which indicated an acidic pH of ca. 2.

The mimics are each di-TFA salts as a result of the deprotection step and purification, but this salt alone should not be acidic. Instead, a small amount of free TFA may remain as a contaminant from either the deprotection or subsequent HPLC purification (with 0.1 % v/v TFA as an eluent). In 100 µL of pure water (approximately the volume of the mimic stocks), just 0.1 µL of residual TFA would lower the pH to 2. Assuming that the small molecule elutes from the HPLC over a one minute period (10 mL), the collected fraction would contain 10 µL TFA, so if 1 % of this amount remained it would explain the low pH

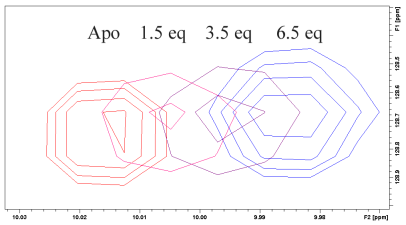
measured. The molecules were lyophilized and left under high vacuum for a prolonged period, so very little TFA (BP 72 °C) should remain, but this was the only stage where acid could have been introduced. The acidity of the lyophilized compounds was neutralized with dilute NaOH to ca. pH 7.0, measured by narrow-range pH paper.

The NMR titrations were repeated with neutralized **1** and heavily buffered (75 mM phosphate) Grb2 SH3C (340 μM, pH 7.2). The addition of the first few equivalents of **1** proceeded without any visible precipitation, but as 3.5 eq were added some precipitation became apparent. The addition of a small amount of concentrated buffer returned this to solution. Ultimately, a maximum of 6.5 eq of **1** were added, resulting in very small shifts, but with a clear dose response. To further analyze the binding behaviour, the change in chemical shift upon addition of ligand was plotted for each residue. The chemical shifts of ^{1}H and ^{15}N nuclei differ in magnitude, so they are averaged using the formula:

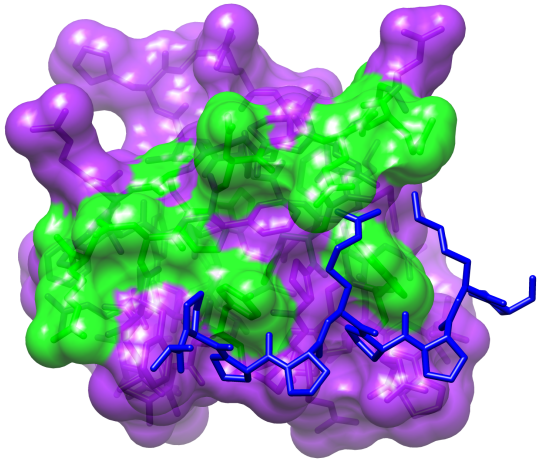
$$Shift = \sqrt{0.5 \times [\delta_H^2 + (\alpha \times \delta_N^2)]}$$

where α is a scale factor, in this case taken to be 0.14.[22] The largest shifts found were extremely small—ca. 0.04 ppm—but are consistent with previous work on Grb2 SH3C, which reported maximal shifts of 0.03 ppm.[24] Setting a cutoff of 0.015 ppm to define interacting residues, twelve were found to interact with mimic **1** (Figure 3-14). These residues are relatively contiguous, and occupy sites along a similar groove as the Gab2a peptide. The mimic is much smaller than the peptide, however, so in reality this indicates relatively non-specific interaction along a broad face of the protein. This is also consistent with the small NMR shifts observed, as the experiment would reveal only the average of all of the potential interactions.

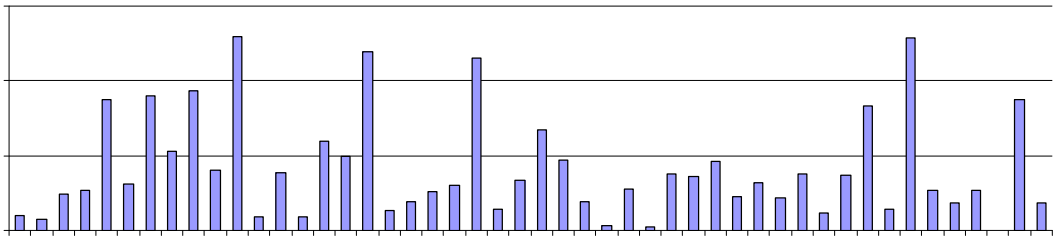
A



C



B



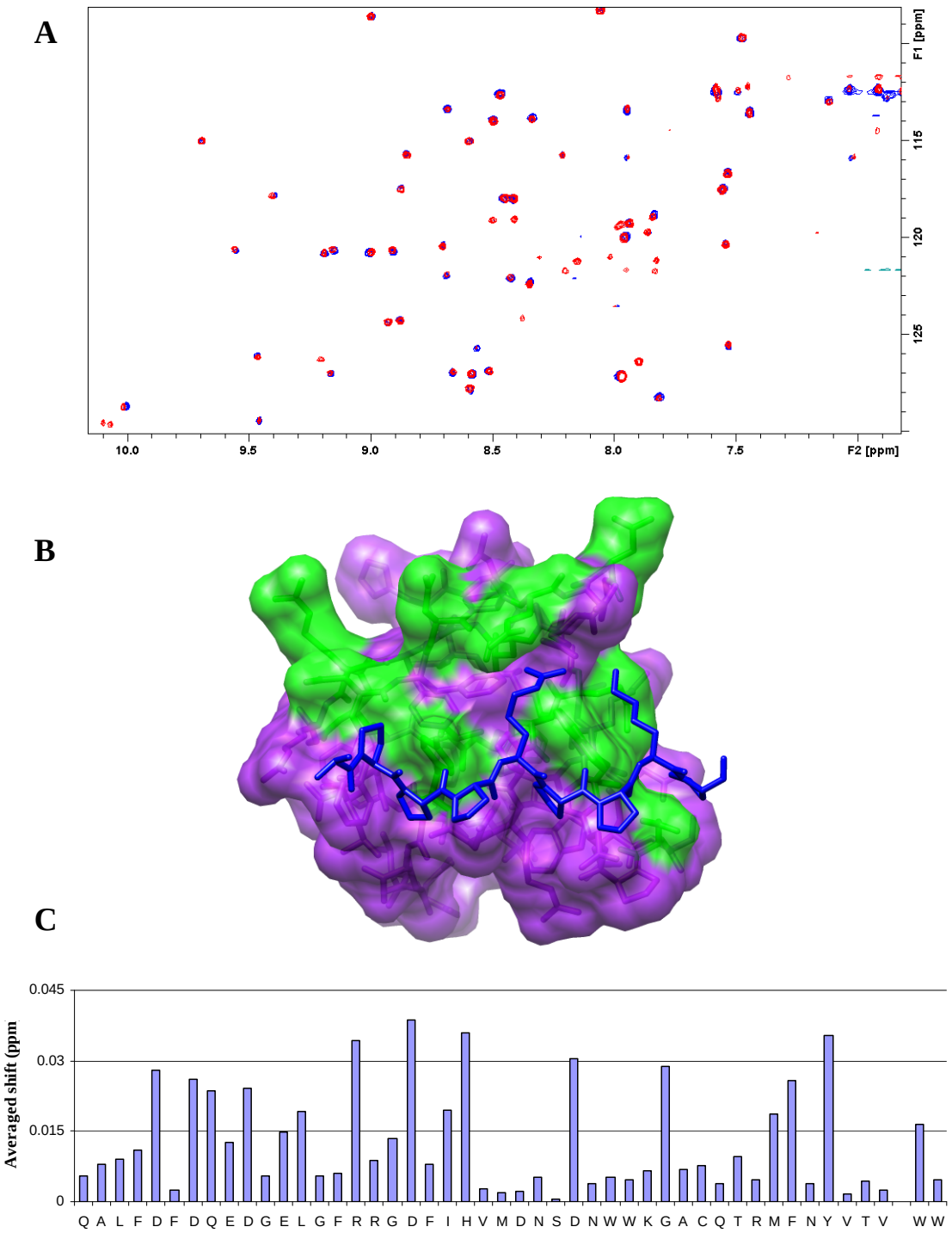


Figure 3-15: (A): ^{15}N HSQC overlay of apo structure (160 μM, pH 7.2; red) and with 10 eq **43** (blue). The most significant effect is the disappearance of contaminant peaks from the complex that are present in the apo structure. (B): model of the interacting residues found by the NMR shift experiments. Residues which shifted > 0.015 ppm are coloured green and mapped onto the crystal structure of Grb2 SH3C binding the Gab2a PPII peptide (2W0Z; blue). Binding occurs over a large surface, incorporating many contiguous residues in several binding sites. (C): histogram of HSQC shifts for each residue in Grb2 SH3C.

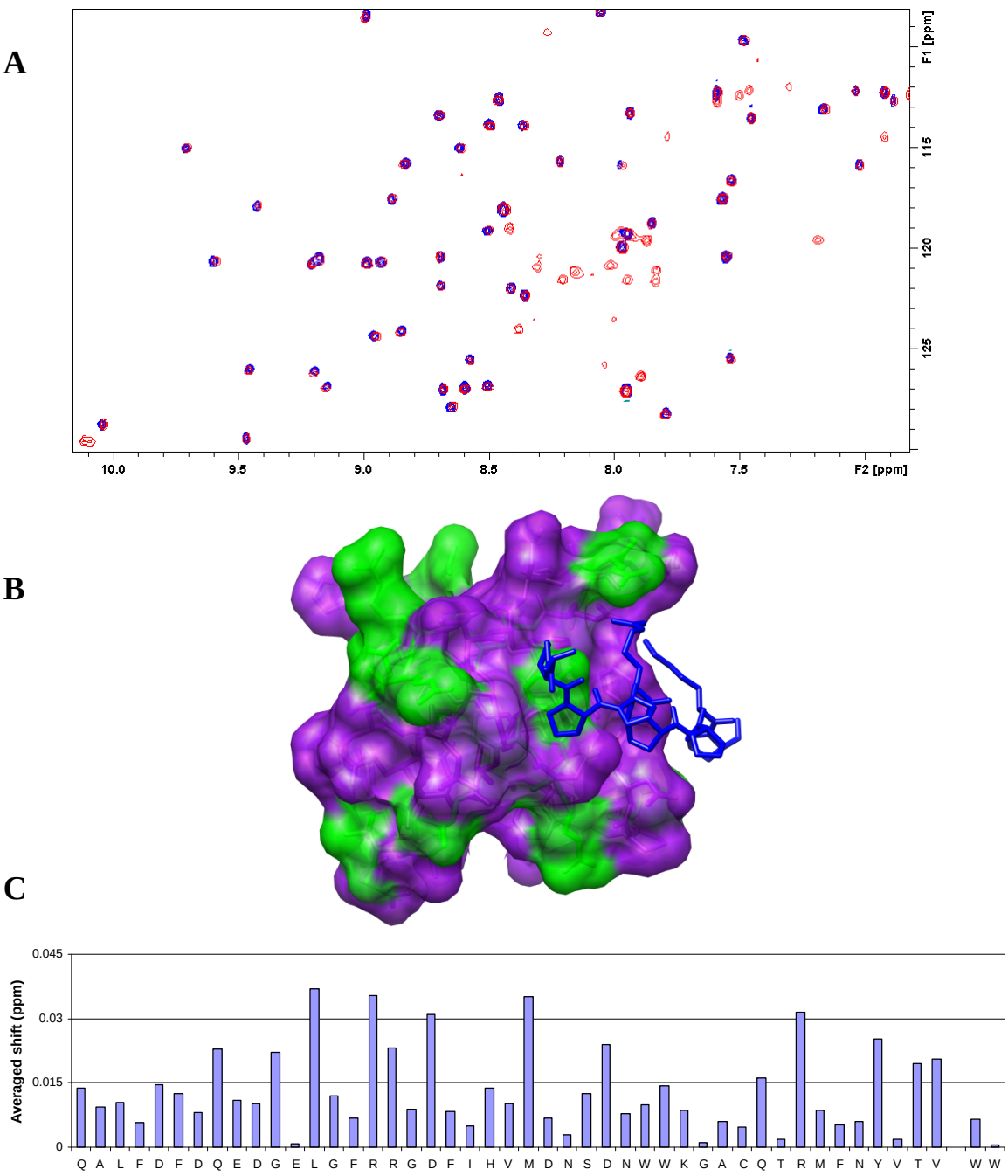


Figure 3-16: (A): ^{15}N HSQC overlay of apo structure (red; pH 7.6; 40 μM) and with 20 eq 42 (blue). The most significant effect is the disappearance of contaminant peaks from the complex that are present in the apo structure. A few genuine shifts are apparent, but most are subtle. (B): model of the interacting residues found by the NMR shift experiments. Residues which shifted > 0.015 ppm are coloured green and mapped onto the crystal structure of Grb2 SH3C binding the Gab2a PPII peptide (2W0Z; blue). There are few contiguous regions and very little overlap with the peptide binding site, indicating nonspecific binding. (C): histogram of HSQC shifts for each residue in Grb2 SH3C.

Finally, titrations were attempted with the extended mimic **42** which showed some activity in the competition assay. This molecule was relatively insoluble, even in 5 % DMSO solutions, and as it was titrated into the protein a precipitate formed. By ^{1}H NMR, it

was clear that a great deal of this precipitate was mimic as the integrated peaks from the mimic relative to protein were much smaller than the appropriate molar ratio. The selective removal of protein contaminant peaks also occurred. To maximize the molar ratio of compound to protein, the protein concentration was reduced to just 40 μM and mimic added in twenty-fold excess. While this resulted in a very clean spectrum, only small resonance shifts were apparent (Figure 3-16). When mapped to the Grb2 SH3C crystal structure, few of these residues are contiguous and there is very little overlap with the peptide binding site. This implies strongly that the binding of **42** is nonspecific.

3.7. Discussion

Computational modelling and structural knowledge of Grb2 SH3C and benzoylureas suggest that a modified form of the scaffold may be able to mimic the PPII helix. Synthesis of the appropriately-functionalized molecules, however, required more careful consideration of protecting groups than had previously been required for this scaffold. Principally, the cationic Arg and Lys sidechains had to be completely masked until after the benzoylurea coupling, and ideally until the final synthetic steps due to the difficulty in handling these highly polar amines and guanidines. An azide proved to be an effective masked primary amine, but its synthesis required the initial formation of a potentially dangerous azidopropanol, which was both volatile and a possible explosive. An improved synthesis would avoid this intermediate, perhaps by instead forming the S_NAr product with bromopropanol (**31**) and displacing the halide with sodium azide. Once this azide was formed and conditions for selectively reducing the aromatic nitro group were found, this sidechain proved robust, tolerant of strong base and moderate acid and bench-stable for a prolonged period. As long as strongly reducing conditions are avoided, azides could prove useful as masked amines in other peptidomimetics.

Ultimately, despite the protecting group manipulation required in synthesizing these benzoylureas, the procedure was optimized to the point of providing 150 mg of mimic **1**. The scaffold also readily incorporated a range of sidechains and can be extended with further urea linkages, although the process requires numerous synthetic steps. The synthesis of large libraries of benzoylureas would be challenging with the solution-phase methods used here, although some progress has recently been reported on the solid-phase synthesis of benzoylureas.[26]

Proof of the biological utility of these molecules, however, can only be obtained experimentally, and here problems arose. SPR measurements indicate that molecule **1** mimicking Arg and Lys sidechains is a relatively potent binder of Grb2, but provide no structural information as to the binding site. The superstoichiometric binding observed at high ligand concentrations, as well as the relative irreproducibility of the binding affinity, hint at complex binding behaviour. This is confirmed by a biochemical competition assay, which fails to show consistent results for the peptidomimetics while still performing well for peptide controls. The further development of this competition assay is complicated by a lack of small molecule positive controls; with only peptides to work with, it is hard to tell if the observed negative results are true negatives, or artifacts due to some specific feature of the benzoylurea ligands.

High-resolution NMR measurements partially explain the anomalous results. While clear dose-response behaviour is observed for mimic **1**, the magnitude of the shifts is small, and much smaller than those observed for the peptide ligands. These small shifts are precedented in the literature for Grb2, but they are nonetheless worrisome as such behaviour is generally characteristic of low occupancy of binding sites or of multipoint interactions. In this case, either or both may be true as at high concentration three of the ligands perturbed residues corresponding to multiple binding sites. At the same time, as there is uncertainty

regarding the affinities of the molecule for any particular site, it is impossible to estimate the occupancy. These issues are compounded by the high concentration of ligand required for NMR experiments and the relatively low solubility of the mimics at these concentrations, which precludes testing higher molar ratios of compound.

Unfortunately, the less than ideal binding of these mimics does not conclusively inform the design of future PPII mimics. It is possible that the computational design of the modified benzoylurea scaffold is correct, and that particular characteristics of Grb2 SH3C make finding any small molecule ligands for the protein difficult. If this is the case, the molecules synthesized here may match alternative RxxK-binding SH3 domains, but this could only be determined by screening those proteins. Alternatively, the design of the molecules may be insufficient. The computational overlay suggests that the polar sidechains are well-positioned, but the overall flexibility of the scaffold and the long sidechains may impart too great an entropic cost for binding to be favourable. Additionally, the aliphatic sidechains tested may not be adequate isosteres of proline, and larger groups could perhaps improve affinity by increased hydrophobic contact. Entirely novel scaffolds may also be needed for mimicking PPII helices, ideally incorporating high aqueous solubility, constrained conformation, and spanning the necessary 9 Å. Such scaffolds could be tested on Grb2 SH3C, or preferably on several PPII-binding proteins, to explore the feasibility of mimicking this peptide structure.

3.8. Conclusion and future work

Reported here is the design and synthesis of some of the first PPII peptidomimetics. These molecules incorporate cationic sidechains, and can be readily functionalized to incorporate a range of sidechains. Biological data from SPR, a competition binding assay, and NMR indicate that several of the putative PPII mimics bind to Grb2 SH3C. While these are likely weak, multipoint interactions, they are among the first molecules reported to bind

to this important signalling domain. Further structural studies, including additional NMR experiments and crystallography, may help better characterize the binding of these molecules. This could guide additional synthetic efforts, using the benzoylurea scaffold or a different structure that might better match the Gab2a peptide. Additionally, selectivity studies with other SH3 domains could help indicate whether these molecules may be useful as chemical probes for the Grb2/Gab2 protein-protein interaction.

3.9. Experimental details: SPR

3.9.1. General SPR methods

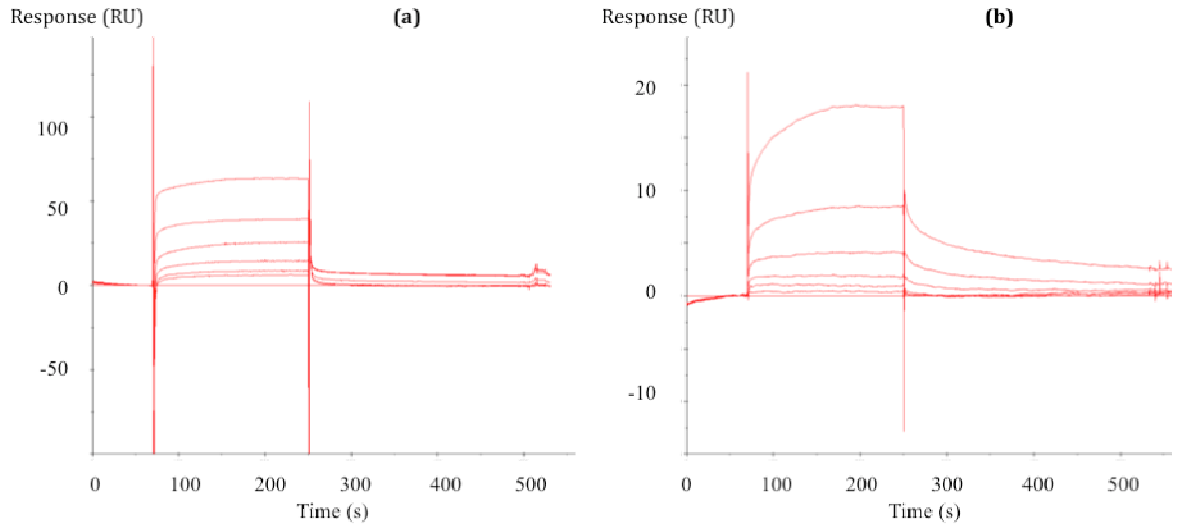
SPR was performed using a Biacore T200 optical biosensor. A fresh CM5 chip was docked to the instrument and hydrated with duplicate 12 s injections of 50 mM NaOH, 10 mM Glycine•HCl pH 1.5, and 0.1 % SDS at 50 μL.min^{-1} flow rate. The flow cells were then normalized with 70 % glycerol using the default normalization wizard. The chip was primed into HBS-N running buffer (10 mM Hepes pH 7.4, 150 mM NaCl), and the flow cells activated with a 1:1 mixture of 0.4 M EDC and 0.1 M NHS for 10 minutes at 10 μL.min^{-1} and 25 ˚C. Grb2 SH3C (100 μM in pH 4.0 acetate buffer) was coupled to the surface. After coupling, all flow cell surfaces were deactivated with 4 x 30 s injections of a 1:1 mixture of 1.0 M ethanolamine and running buffer. The chip was subsequently primed twice into the experimental running buffer. Molecules were tested at a flow rate of 40 μL.min^{-1} at 25 ˚C, with a data collection rate of 10 Hz. The data were analyzed using standard double referencing procedures by subtracting the response from the protein-free cell and the response from a blank injection of running buffer from the response from the SH3C flow cell.[27]

3.9.2. SPR measurements of 1st generation mimics

Untagged and unlabelled Grb2 SH3C was bound to the flow cell surface to a level of 1400 RU. For the binding screen, each molecule was dissolved in running buffer (HBS-N, 5.0 % DMSO, 0.01 % Tween 20) at 50 μM and 100 μM concentrations. The most successful binder **1** was then serially diluted two-fold from a concentration of 100 μM to a concentration of 3.125 μM in running buffer not containing DMSO (HBS-N + 0.01 % Tween 20). Data was collected according to the general SPR procedure and fitted to a one-to-one binding model using nonlinear regression analysis in GraphPad Prism software.

3.9.3. SPR measurements of 2nd generation mimics

Untagged an unlabelled Grb2 SH3C was covalently bound to a level of 1100 RU. On a separate flowcell, ^{15}N labeled Grb2 SH3C was covalently bound to the same level. Mimics were dissolved either in HBS-N buffer + 0.01 % Tween 20, or in a similar buffer containing 5 % DMSO, depending on solubility. Serial two-fold dilutions were performed from a concentration of 200 μM to a final concentration of 6.25 μM. Data was collected according to the general SPR procedure and fitted to a one-to-one binding model using nonlinear regression analysis in GraphPad Prism software.



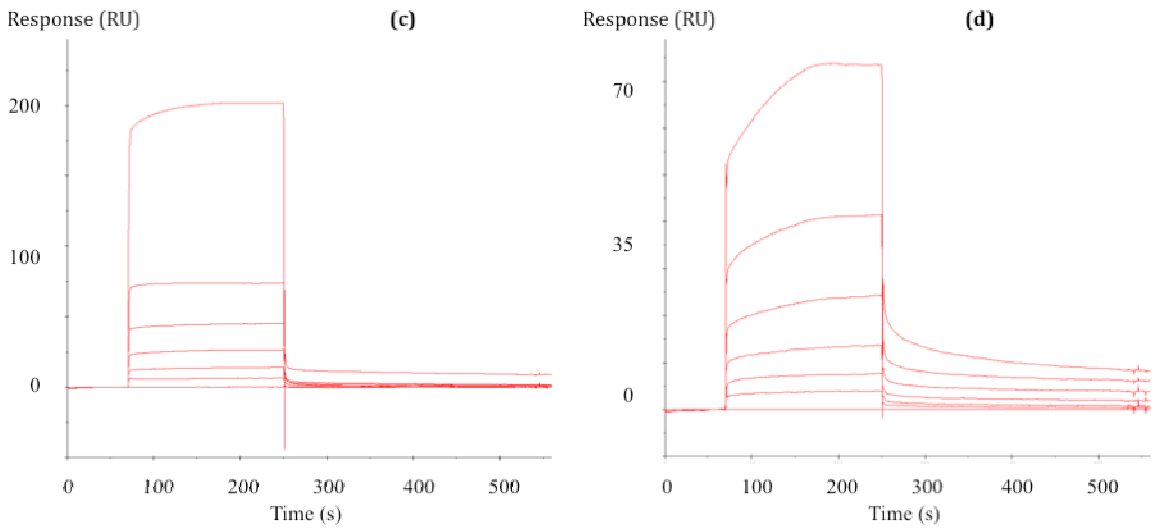


Figure 3-17: Sensograms of 2nd generation mimics binding to ^{15}N-labeled Grb2 SH3C. (a) **37** (b) **38** (c) **39** (d) **41**. Mimic concentrations are a two-fold dilution series from 200 µM (highest response) to 6.25 µM (lowest non-zero response).

3.10. Experimental details: competition binding assay

GST–Grb2 SH3C (10 µg per sample) was pre-incubated with each inhibitor (2.0 mM) in duplicate for 1 hour at 4 ˚C in assay buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT, 0.05 % Tween 20, 5 % DMSO). Biotinylated Gab2b peptide (binding affinity 3.2 µM for Grb2) was synthesized as previously described.[25] 0.3 nmol of peptide per µL of packed streptavadin-sepharose beads were incubated together for 30 minutes in assay buffer before extensive washing of the beads to remove any unbound peptide. Next, 8 µL of peptide-coated beads were transferred to each assay tube containing GST–Grb2 SH3C and compound for 30 minutes, the solution clarified by centrifugation (2000 x *g* for 1 minute) and the beads washed three times with PBS-T (phosphate-buffered saline, 0.05 % Tween 20). The samples were analyzed by SDS-PAGE and visualized after staining the protein bands with Coomassie Blue dye. As controls, parallel samples were run in the absence of inhibitor, and with preincubation of ten-fold or 100-fold molar excess of free, unbiotinylated Gab2b peptide. The band intensities, indicating the levels of GST–Grb2

SH3C binding to the beads, were quantified by densitometry and scaled relative to the control sample without inhibitor (100 % level).

For concentration-response experiments with **42**, the above procedure was followed with increasing concentrations of the mimic.

3.11. Experimental details: NMR

All spectra were collected using a Bruker AVII 700 spectrometer (fitted with an inverse TCI cryoprobe optimized for ^{1}H observation and running Topspin 2.1 software; Bruker, Germany). Experiments were performed at 25 ˚C. Resonances are reported in ppm relative to D_2O (δ_H = 4.72 ppm). The instrument was locked to the deuterium signal, and the water signal was suppressed by presaturating its resonance. Grb2 SH3C was uniformly ^{15}N isotopically labeled and dissolved in phosphate buffer (90:10 H_2O:D_2O, 35 mM phosphate pH 7.2 + 45 mM NaCl + 3.0 mM DTT). A 3 mM NMR tube was charged with 160 μL of sample and the ^{15}N HSQC spectrum recorded. Resonances were compared to the literature values.[24] For titration experiments, varying molar equivalents of ligand were added to the protein solution described above. Data was analyzed using Bruker's Topspin 2.1 software.

3.12. Experimental details: synthesis

3.12.1. Solvents and Reagents

All non-aqueous reactions were carried out under an atmosphere of argon or nitrogen in oven or flame-dried glassware unless otherwise stated. Anhydrous THF and DCM (from commercial sources) were dried on an MB-SPS-800 solvent purification system. All other solvents and reagents were used directly from commercial suppliers without further purification. Inorganic solutions refer to saturated aqueous solutions, unless otherwise indicated. Brine refers to a saturated aqueous solution of sodium chloride. pH 7 buffer was

prepared by dissolving KH_2PO_4 (85 g) and NaOH (14.5 g) in 950 mL of water. Mixtures of solvents are quoted as ratios of volume:volume.

3.12.2. Chromatography

Flash column chromatography was carried out using commercial Kieselgel 60 silica gel (40-63 μm). Thin-layer chromatography was carried out using Merck Kieselgel 60 F254 (230-400 mesh) fluorescent treated silica, visualized under UV light (254 nm) and by staining with aqueous potassium permanganate or ninhydrin solution. High performance liquid chromatography was performed using a 1525 pump, 2707 autosampler, and 2849 detector, all from Waters. Phenomonex Luna columns (250 mm long, 5 μm beads, C18 reverse-phase medium) were used for HPLC separations. Analytical HPLC was run using 1 mL/min flow through a 4.6 mm-diameter column. Sample injections for analytical runs consisted of 40 μL of a 1 $mg.mL^{-1}$ sample solution. Semi-preparative HPLC was run using 10 $mL.min^{-1}$ flow through 21.2 mm-diameter column. Sample injections for semi-preparative runs consisted of 500 μL of solution containing no more than 50 mg of sample. HPLC solvents were degassed by sonication for 30 minutes and contained 0.1 % v/v TFA.

3.12.3. Infrared Spectroscopy

Infrared spectra were recorded on a Bruker Tensor 27 Fourier transform-IR spectrometer from a thin film deposited onto a diamond ATR module. Only selected maximum absorbances (υ_{max}) of the most intense peaks are reported (cm^{-1}).

3.12.4. Mass Spectrometry

Low-resolution mass spectra were recorded on a Waters LCT premier XE Micromass spectrometer under conditions of electrospray ionization (ESI). High-resolution mass spectra were recorded on a Bruker MicroTof mass spectrometer (ESI, electron impact [EI],

or field ionization [FI]) by the internal service at the Department of Chemistry, University of Oxford. Values reported are the ratio of mass to charge in Daltons.

3.12.5. Nuclear Magnetic Resonance Spectroscopy

NMR spectra were recorded on a Bruker DRX500 (^{1}H: 500 MHz; ^{13}C: 125 MHz), AVC500 (^{1}H: 500 MHz; ^{13}C: 125 MHz), DPX400 (^{1}H: 400 MHz; ^{13}C: 100 MHz), AVIII 400 HD Nanobay (^{1}H: 400 MHz; ^{13}C: 100 MHz), DPX300 (^{1}H: 300 MHz; ^{13}C: 75 MHz), DPX250 (^{1}H: 250 MHz; ^{13}C: 62.5 MHz), or DPX200 (^{1}H: 200 MHz; ^{13}C: 50 MHz). Chemical shifts are quoted relative to the residual non-deuterated solvent peak. ^{1}H spectra are reported as follows: chemical shift δ/ppm (number of protons, multiplicity, coupling constant J/Hz [where appropriate], assignment). Multiplicity is abbreviated as follows: s, singlet; br, broad; d, doublet; t, triplet; q, quartet; quint, quintet, sept, septet; m, multiplet. Peaks which could not be assigned in the ^{1}H spectra due to the similarity of aromatic protons are indicated by (Ar). Coupling constants *J* are reported to the nearest 0.1 Hz as observed. ^{13}C spectra are reported in δ/ppm.

3.12.6. Melting Point

Melting points were determined using a Leica Galen III hot-stage microscope apparatus and are reported uncorrected in degrees Celsius (˚C).

3.12.7. Polarimetry

Optical rotations were recorded on a Perkin-Elmer 241 polarimeter with a path length of 1 dm and are reported in deg $dm^{-1}.cm^{3}.g^{-1}$ using concentrations in g.100 ml^{-1}.

3.12.8. Naming and Numbering of Compounds

Compound names are those generated from ChemBioDraw Ultra 12.0 following IUPAC conventions. Numbers within names refer to the IUPAC system of numbering, also from

ChemBioDraw. When assigning NMR spectra, a separate atom numbering scheme is used as indicated on the structure. This numbering does not follow any particular convention.

3.13. References

(1) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. Host-guest scale of left-handed polyproline II helix formation. *Proteins Struct. Funct. Bioinforma.* **2003**, *53*, 68–75.

(2) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9190–9195.

(3) Rucker, A. L.; Creamer, T. P. Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Sci. Publ. Protein Soc.* **2002**, *11*, 980–985.

(4) Liu, Z.; Chen, K.; Ng, A.; Shi, Z.; Woody, R. W.; Kallenbach, N. R. Solvent Dependence of PII Conformation in Model Alanine Peptides. *J. Am. Chem. Soc.* **2004**, *126*, 15141–15150.

(5) Whittington, S. J.; Creamer, T. P. Salt Bridges Do Not Stabilize Polyproline II Helices. *Biochemistry (Mosc.)* **2003**, *42*, 14690–14695.

(6) Petrella, E. C.; Machesky, L. M.; Kaiser, D. A.; Pollard, T. D. Structural Requirements and Thermodynamics of the Interaction of Proline Peptides with Profilin. *Biochemistry (Mosc.)* **1996**, *35*, 16535–16543.

(7) Ude, S.; Lassak, J.; Starosta, A. L.; Kraxenberger, T.; Wilson, D. N.; Jung, K. Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. *Science* **2013**, *339*, 82–85.

(8) Doerfel, L. K.; Wohlgemuth, I.; Kothe, C.; Peske, F.; Urlaub, H.; Rodnina, M. V. EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. *Science* **2013**, *339*, 85–88.

(9) Fillon, Y. A.; Anderson, J. P.; Chmielewski, J. Cell Penetrating Agents Based on a Polyproline Helix Scaffold. *J. Am. Chem. Soc.* **2005**, *127*, 11798–11803.

(10) Zhang, R.; Nickl, C. K.; Mamai, A.; Flemer, S.; Natarajan, A.; Dostmann, W. R.; Madalengoitia, J. s. Poly-l-proline type II peptide mimics as probes of the active site occupancy requirements of cGMP-dependent protein kinase. *J. Pept. Res.* **2005**, *66*, 151–159.

(11) Zaminer, J.; Brockmann, C.; Huy, P.; Opitz, R.; Reuter, C.; Beyermann, M.; Freund, C.; Müller, M.; Oschkinat, H.; Kühne, R.; Schmalz, H. Addressing Protein-Protein Interactions with Small Molecules: A Pro-Pro Dipeptide Mimic with a PPII Helix Conformation as a Module for the Synthesis of PRD-Binding Ligands. *Angew. Chem. Int. Ed.* **2010**, *49*, 7111–7115.

(12) Tremmel, P.; Geyer, A. An Oligomeric Ser-Pro Dipeptide Mimetic Assuming the Polyproline II Helix Conformation. *J. Am. Chem. Soc.* **2002**, *124*, 8548–8549.

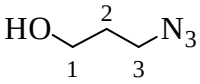
(13) Raghavan, B.; Skoblenick, K. J.; Bhagwanth, S.; Argintaru, N.; Mishra, R. K.; Johnson, R. L. Allosteric Modulation of the Dopamine D2 Receptor by Pro-Leu-Gly-NH2 Peptidomimetics Constrained in Either a Polyproline II Helix or a Type II β-Turn Conformation. *J. Med. Chem.* **2009**, *52*, 2043–2051.

(14) Bhagwanth, S.; Mishra, R. K.; Johnson, R. L. Development of peptidomimetic ligands of Pro-Leu-Gly-NH $_2$ as allosteric modulators of the dopamine D $_2$ receptor. *Beilstein J. Org. Chem.* **2013**, *9*, 204–214.

(15) Zhang, R.; Madalengoitia, J. S. Design, Synthesis and Evaluation of Poly-l-Proline Type-II Peptide Mimics Based on the 3-Azabicyclo[3.1.0]hexane System. *J. Org. Chem.* **1999**, *64*, 330–331.
(16) Rodriguez, J. M.; Hamilton, A. D. Benzoylurea Oligomers: Synthetic Foldamers That Mimic Extended α Helices. *Angew. Chem. Int. Ed.* **2007**, *46*, 8614–8617.
(17) Rodriguez, J. M.; Ross, N. T.; Katt, W. P.; Dhar, D.; Lee, G.; Hamilton, A. D. Structure and Function of Benzoylurea-Derived α-Helix Mimetics Targeting the Bcl-xL/Bak Binding Interface. *Chem. Med. Chem.* **2009**, *4*, 649–656.
(18) Thompson, S.; Hamilton, A. D. Amphiphilic α-helix mimetics based on a benzoylurea scaffold. *Org. Biomol. Chem.* **2012**, *10*, 5780.
(19) Stewart, J. J. P. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *J. Mol. Model.* **2004**, *10*, 155–164.
(20) Bald, E.; Saigo, K.; Mukaiyama, T. A Facile Synthesis of Carboxamides by Using 1-Methyl-2-Halopyridinium Iodides as Coupling Reagents. *Chem. Lett.* **1975**, *4*, 1163–1166.
(21) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; illustrated edition.; University Science, 2005.
(22) Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
(23) Alberty, R. A.; Hammes, G. G. Application of the Theory of Diffusion-controlled Reactions to Enzyme Kinetics. *J. Phys. Chem.* **1958**, *62*, 154–159.
(24) Lin, C.-C.; Melo, F. A.; Ghosh, R.; Suen, K. M.; Stagg, L. J.; Kirkpatrick, J.; Arold, S. T.; Ahmed, Z.; Ladbury, J. E. Inhibition of Basal FGF Receptor Signaling by Dimeric Grb2. *Cell* **2012**, *149*, 1514–1524.
(25) Harkiolaki, M.; Tsirka, T.; Lewitzky, M.; Simister, P. C.; Joshi, D.; Bird, L. E.; Jones, E. Y.; O'Reilly, N.; Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the SH3C Domain of Grb2. *Structure* **2009**, *17*, 809–822.
(26) Douat-Casassus, C.; Pulka, K.; Claudon, P.; Guichard, G. Microwave-Enhanced Solid-Phase Synthesis of N,N′-Linked Aliphatic Oligoureas and Related Hybrids. *Org. Lett.* **2012**, *14*, 3130–3133.
(27) Myszka, D. G. Improving biosensor analysis. *J. Mol. Recognit.* **1999**, *12*, 279–284.
(28) Badiang, J. G.; Aubé, J. One-Step Conversion of Aldehydes to Oxazolines and 5,6-Dihydro-4H-1,3-oxazines Using 1,2- and 1,3-Azido Alcohols. *J. Org. Chem.* **1996**, *61*, 2484–2487.
(29) Shaginian, A.; Whitby, L. R.; Hong, S.; Hwang, I.; Farooqi, B.; Searcey, M.; Chen, J.; Vogt, P. K.; Boger, D. L. Design, Synthesis, and Evaluation of an α-Helix Mimetic Library Targeting Protein−Protein Interactions. *J. Am. Chem. Soc.* **2009**, *131*, 5564–5572.

3.14. Experimental Data for Synthesized Compounds

3-Azidopropan-1-ol (5)

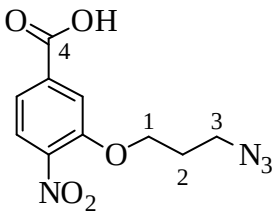


Sodium azide (1.95 g, 30 mmol, 1.05 eq) was dissolved in DMSO (60 mL). To this solution was added 3-bromo-1-propanol (3.97 g, 28.6 mmol, 1 eq) and the mixture heated to 70 ºC and stirred overnight. The reaction was partitioned between water (100 mL) and ether (100 mL) and washed with ether (3 x 60 mL). The organic layers were combined, washed with brine, dried ($MgSO_4$), filtered, and concentrated at 500 mBar and 40 ºC. Purification *via* flash column chromatography (1:1 ether:petrol) yielded the title compound as a volatile, colourless oil (2.70 g, 26.7 mmol, 93 %): δ_H (500 MHz, $CDCl_3$) 3.82 (2H, q, *J* 5.7, H1), 3.53 (2H, quint, *J* 3.5, H3), 1.90 (2H, quint, *J* 6.3, H2). δ_C (100 MHz, $CDCl_3$) 60.0 (C1), 48.5 (C3), 31.4 (C2). IR υ_{max} 3335, 2946, 2882, 2090, 1045. HRMS (FI) found 101.0590. $C_3H_7N_3O^+$ $[M]^+$ requires 101.0590. Spectroscopic measurements agreed with literature values.[26]

3.14.1. General Procedure for S_NAr Coupling

Following the procedure of Boger *et al*,[27] alcohol (1.2 eq) was added dropwise to a stirred solution of sodium hydride (2.5 eq) in THF at 0 ºC. The mixture was stirred for 15 minutes at 0 ºC before 3-fluoro-4-nitrobenzoic acid **6** (1.00 g, 5.4 mmol, 1.0 eq) was added. The reaction was stirred for 5 minutes at 0 ºC, then 2 hours at room temperature, before being diluted with EtOAc (20 mL). NH_4Cl (20 mL) was added, and the mixture was washed with HCl (0.5 M, 20 mL x 3). The organic layer was collected, dried ($MgSO_4$), filtered, and concentrated, and the product purified *via* flash column chromatography.

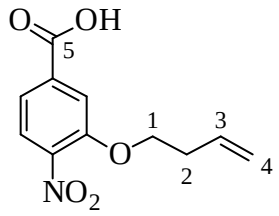
3-(3-Azidopropoxy)-4-nitrobenzoic acid (7)



Following the general procedure for S_NAr coupling, azide **5** (655 mg, 6.5 mmol, 1.2 eq) was added to acid **6** and purified (3:2

EtOAc:Pet + 1 % AcOH) to yield the title compound as a yellow solid (1.40 g, 5.26 mmol, 97 %): MP 123-126. δ_H (400 MHz, MeOD) 7.86 (1H, d, *J* 10.9, Ar), 7.85 (1H, s, Ar), 7.72 (1H, dd, *J* 8.3; 1.6, Ar), 4.30 (2H, t, *J* 5.9, H1), 3.56 (2H, t, *J* 6.6, H3), 2.09 (2H, quint, *J* 6.2, H2). δ_C (125 MHz, MeOD) 166.6 (C4), 151.6 (Ar), 143.0 (Ar), 136.0 (Ar), 125.1 (Ar), 121.8 (Ar), 115.6 (Ar), 66.6 (C1), 28.5 (C2). IR υ_{max} 3061, 2924, 2853, 2102, 1690, 1589, 1530, 1259. HRMS (ESI) found 265.0569 $C_{10}H_9N_4O_5^-$ $[M-H]^-$ requires 265.0578.

3-(But-3-en-1-yloxy)-4-nitrobenzoic acid (8)



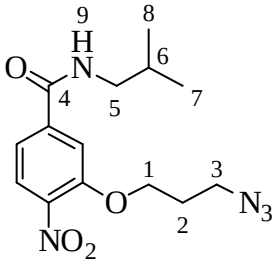
Following the general procedure for S_NAr coupling, 3-buten-1-ol (584 mg, 8.1 mmol, 1.5 eq) was added to acid **6** and purified (30:1 DCM:MeOH + 1% AcOH) to yield the title compound as a yellow solid (1.055 g, 4.45 mmol, 82 %): δ_H (250 MHz, MeOD) 7.79-7.73 (2H, m, Ar), 7.64 (1H, dd, *J* 7.6; 1.6, Ar), 5.98-5.80 (1H, m, H3), 5.20-5.04 (2H, m, H4), 4.19 (2H, t, *J* 4.2, H1), 2.54 (2H, qt, *J* 6.6; 1.5, H2). δ_C (62.5 MHz, MeOD) 166.7 (C5), 151.7 (Ar), 143.0 (Ar), 135.8 (Ar), 134.0 (C3), 125.0 (Ar), 121.7 (Ar), 117.1 (C4), 115.8 (Ar), 69.3 (C1), 33.4 (C2). IR υ_{max} 2950, 1687, 1527, 1251 HRMS (ESI) found 260.0524 $C_{11}H_{11}NO_5Na^+$ $[M+Na]^+$ requires 260.0529.

3.14.2. General Procedure I for Amidation

Oxalyl chloride (10 eq) was added to a solution of the substituted nitrobenzoic acid (1.0 eq) dissolved in DCM (5 mL/mmol). DMF (~0.5 eq) was added slowly and the mixture stirred vigorously for 1 hour. Solvent was removed under reduced pressure, and the resulting oil azeotroped three times with chloroform. The oil was dissolved in DCM (5 mL/mmol) and cooled to 0 ºC. DIPEA (1.5 eq) was added to the solution dropwise, followed by the amine (1.3 eq) added dropwise. The solution was allowed to warm to room temperature and stirred overnight. Methanol (0.5 mL) was added, and the mixture was washed with NH_4Cl (10

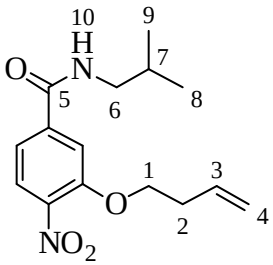
mL/100 mg), and extracted with DCM (3 x 10 mL/100 mg). The organic layers were combined, dried ($MgSO_4$), filtered, and concentrated, and the product purified *via* flash column chromatography.

3-(3-Azidopropoxy)-*N*-isobutyl-4-nitrobenzamide (9)



Following the general procedure I for amidation, azido benzoic acid **7** (100 mg, 0.38 mmol, 1 eq) was coupled to isobutyl amine (36 mg, 0.49 mol, 1.3 eq) and purified (2:1 ether:petrol) to yield the title compound as a viscous orange oil (120 mg, 0.37 mmol, 99 %): δ_H (400 MHz, $CDCl_3$) 7.86 (1H, d, *J* 8.3, Ar), 7.61 (1H, d, *J* 1.6, Ar), 7.28 (1H, dd, *J* 8.1; 1.6, Ar), 6.26 (1H, s, H9), 4.27 (2H, t, *J* 5.8, H1), 3.58 (2H, t, *J* 6.4, H3), 3.30 (2H, t, *J* 6.7, H5), 2.10 (2H, quint, *J* 6.1, H2), 1.92 (1H, sept, *J* 1.9, H6), 0.99 (6H, d, *J* 6.7, H7 and H8). δ_C (62.5 MHz, $CDCl_3$) 165.4 (C4), 152.6 (Ar), 141.6 (Ar), 140.6 (Ar), 126.2 (Ar), 118.1 (Ar), 114.5 (Ar), 66.6 (C1), 48.1 (C3), 48.0 (C6), 29.7 (C6), 28.5 (C2), 20.6 (C7 and C8). IR υ_{max} 3314, 3083, 2959, 2100, 1645, 1527, 1254. HRMS (ESI) found 344.1320 $C_{14}H_{19}N_5O_4Na^+$ $[M+Na]^+$ requires 344.1329.

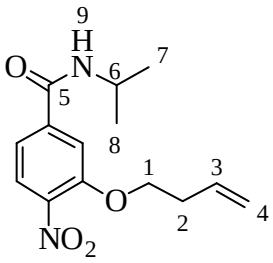
3-(But-3-en-1-yloxy)-*N*-isobutyl-4-nitrobenzamide (10)



Following the general procedure I for amidation, alkene benzoic acid **8** (1.04 g, 4.4 mmol, 1 eq) was coupled to isobutyl amine (418 mg, 0.74 mol, 1.3 eq) and purified (1:1 ether:petrol) to yield the title compound as a viscous yellow oil (1.17 g, 4.01 mmol, 91 %): δ_H (250 MHz, $CDCl_3$) 7.80 (1H, d, *J* 8.3, Ar), 7.58 (1H, d, *J* 1.5, Ar), 7.29 (1H, dd, *J* 8.3; 1.6, Ar), 6.59 (1H, t, *J* 4.9, H10), 5.98-5.79 (1H, m, H3), 5.23-5.09 (2H, m, H4), 4.19 (2H, t, *J* 6.6, H1), 3.28 (2H, t, *J* 6.5, H6), 2.58 (2H, q, *J* 6.7, H2), 1.92 (1H, sept, *J* 6.7, H7), 0.98 (6H, d, *J* 6.7, H8 and H9). δ_C (62.5 MHz, $CDCl_3$) 166.0 (C5), 152.7 (Ar), 141.7 (Ar), 140.5

(Ar), 133.8 (C3), 125.9 (Ar), 118.3 (Ar), 118.0 (C4), 114.5 (Ar), 69.6 (C1), 48.0 (C6), 33.6 (C2), 29.0 (C7), 20.6 (C8 and C9). IR υ_{max} 3313, 3081, 2960, 1644, 1529, 1252. HRMS (ESI) found 315.1311 $C_{15}H_{20}N_2O_4Na^+$ $[M+Na]^+$ requires 315.1315.

3-(But-3-en-1-yloxy)-*N*-isopropyl-4-nitrobenzamide (11)

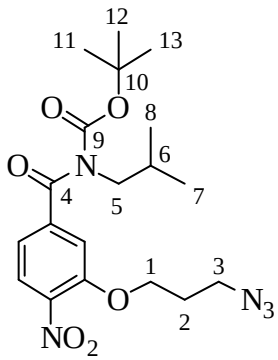


Following the general procedure I for amidation, alkene benzoic acid **8** (75 mg, 0.32 mmol, 1eq) was coupled to isopropyle amine (30 mg, 0.51 mmol, 1.6 eq) and purified (3:2 ether:petrol) to yield the title compound as a viscous yellow oil (50 mg, 0.18 mmol, 57 %): δ_H (250 MHz, $CDCl_3$) 7.80 (1H, d, *J* 8.3, Ar), 7.58 (1H, d, *J* 1.3, Ar), 7.28 (1H, dd, *J* 8.4; 1.5, Ar), 6.24 (1H, d, *J* 6.7, H9), 6.00-5.80 (1H, m, H3), 5.25-5.09 (2H, m, H4), 4.20 (2H, t, *J* 6.6, H1), 2.59 (2H, q, *J* 6.6, H2), 1.29 (6H, d, *J* 6.6, H7 and H8). δ_C (62.5 MHz, $CDCl_3$) 165.1 (C5), 152.7 (Ar), 141.7 (Ar), 140.5 (Ar), 133.8 (C3), 125.9 (Ar), 118.3 (Ar), 118.0 (C4), 114.5 (Ar), 69.6 (C1), 42.8 (C6), 33.6 (C2), 23.1 (C7 and C8). IR υ_{max} 3300, 3079, 2974, 1729, 1641, 1528, 1249. HRMS (ESI) found 279.1338 $C_{14}H_{19}N_2O_4^+$ $[M+H]^+$ requires 279.1339.

3.14.3. General Procedure for the Boc Protection of Amides

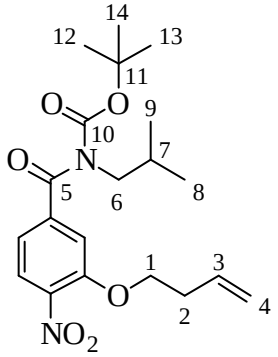
Di-*tert*-butyl dicarbonate (2.0 eq) was added to a stirred solution of the amide (1.0 eq) in DCM. DIPEA (3 eq) and DMAP (0.1 eq) were added and the reaction stirred at room temperature for 2-5 days. The mixture was then washed with NH_4Cl and extracted three times with DCM. The organic layers were combined, dried ($MgSO_4$), filtered, and concentrated, and the product purified *via* flash column chromatography.

***tert*-Butyl (3-(3-azidopropoxy)-4-nitrobenzoyl)(isobutyl)carbamate (12)**



Following the general procedure for the Boc protection of amides, the azide **9** (625 mg, 1.95 mmol, 1 eq) was stirred for 4 days and purified (1:1 ether:petrol) to yield the title compound as a yellow oil (751 mg, 1.78 mmol, 91 %): δ_H (250 MHz, $CDCl_3$) 7.88 (1H, d, *J* 8.3, Ar), 7.26 (1H, d, *J* 1.4, Ar), 7.11 (1H, dd, *J* 8.3; 1.1, Ar), 4.24 (2H, t, *J* 5.7, H1), 3.68 (2H, d, *J* 7.4, H5), 3.61 (2H, t, *J* 6.4, H3), 2.22-2.03 (3H, m, H2 and H6), 1.28 (9H, s, H11-13), 0.99 (6H, d, *J* 6.7, H7 and H8). δ_C (62.5 MHz, $CDCl_3$) 171.3 (C4), 153.7 (C9), 152.4 (Ar), 144.0 (Ar), 140.8 (Ar), 125.7 (Ar), 119.1 (Ar), 113.7 (Ar), 84.3 (C10), 66.6 (C1), 53.1 (C5), 48.1 (C3), 28.8 (C6), 28.5 (C2), 27.9 (C11-C13), 20.6 (C7 and C8). IR υ_{max} 3332, 2960, 2099, 1735, 1673, 1526, 1146. HRMS (ESI) found 444.1843 $C_{19}H_{27}N_5O_6Na^+$ $[M+Na]^+$ requires 444.1854.

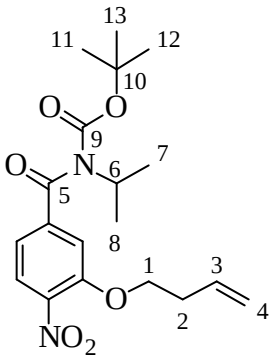
***tert*-Butyl (3-(but-3-en-1-yloxy)-4-nitrobenzoyl)(isobutyl)carbamate (13)**



Following the general procedure for the Boc protection of amides, the alkene **10** (250 mg, 0.87 mmol, 1 eq) was stirred for 2 days and purified (8:1 pet:ether) to yield the title compound as a yellow oil (249 mg, 0.63 mmol, 73 %): δ_H (250 MHz, $CDCl_3$) 7.82 (1H, d, *J* 8.3, Ar), 7.21 (1H, d, *J* 1.5, Ar), 7.07 (1H, dd, *J* 8.3; 1.6, Ar), 5.98-5.79 (1H, m, H3), 5.23-5.09 (2H, m, H4), 4.16 (2H, t, *J* 6.6, H1), 3.65 (2H, d, *J* 7.4, H6), 2.59 (2H, q, *J* 6.7, H2), 2.10 (1H, sept, *J* 6.9, H7), 1.24 (9H, s, H12-14), 0.97 (6H, d, *J* 6.8, H8 and H9). δ_C (62.5 MHz, $CDCl_3$) 171.5 (C5), 153.8 (Ar), 152.6 (C10), 143.8 (Ar), 141.1 (Ar), 133.7 (C3), 125.6 (Ar), 118.9 (Ar), 118.3 (C4), 113.9 (Ar), 84.2 (C11), 69.7 (C1), 53.0 (C6), 33.6 (C2), 28.5 (C7), 27.9 (C12-14), 20.6 (C8 and C9). IR

υ_{max} 3081, 2963, 1737, 1676, 1529, 1273, 1149. HRMS (ESI) found 415.1828 $C_{20}H_{28}N_2O_6Na^+$ $[M+Na]^+$ requires 415.1840.

***tert*-Butyl (3-(but-3-en-1-yloxy)-4-nitrobenzoyl)(isopropyl)carbamate (14)**



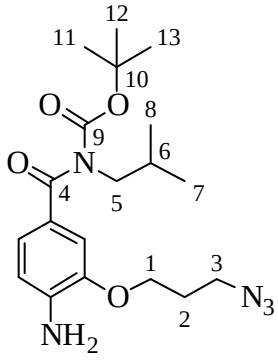
Following the general procedure for the Boc protection of amides, the alkene **11** (50 mg, 0.18 mmol, 1 eq) was stirred for 2 days and purified (8:1 petrol:ether) to yield the title compound as a yellow oil (15 mg, 0.04 mmol, 22 %): δ_H (250 MHz, $CDCl_3$) 7.85 (1H, d, *J* 8.4, Ar), 7.28 (1H, s, Ar), 7.14 (1H, dd, *J* 8.3; 1.4, Ar), 6.03-5.82 (1H, m, H3), 5.27-5.12 (2H, m, H4), 4.29-4.15 (2H, m, H1), 2.63 (2H, q, *J* 6.5, H2), 1.57-1.51 (1H, m, H6), 1.46 (6H, d, *J* 6.9, H7 and H8), 1.25 (9H, s, H11-13). δ_C (62.5 MHz, $CDCl_3$) 171.6 (C5), 153.3 (C9), 152.6 (Ar), 144.1 (Ar), 141.3 (Ar), 133.7 (C3), 125.7 (Ar), 119.0 (Ar), 118.3 (C4), 114.0 (Ar), 84.1 (C10), 69.7 (C1), 53.2 (C6), 33.6 (C2), 27.9 (C7 and C8), 20.8 (C11-13). IR υ_{max} 2976, 2921, 1731, 1676, 1529, 1240. HRMS (ESI) found 401.1667 $C_{19}H_{26}N_2O_6Na^+$ $[M+Na]^+$ requires 401.1683.

3.14.4. General Procedure for Nitro Reduction

Iron powder (10 eq) and ammonium chloride (0.5 eq) were added to a stirred solution of an aromatic nitro compound (1.0 eq) dissolved in ethanol/water (4:1). The mixture was heated to 80 ºC for 30 minutes, then allowed to cool and filtered through Celite® with EtOAc. Solvent was removed *in vacuo* and the product purified *via* flash column chromatography.

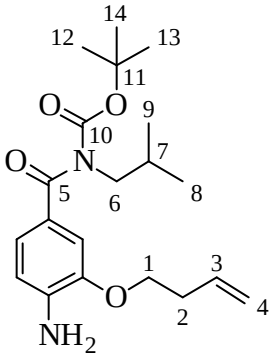
***tert*-Butyl (4-amino-3-(3-azidopropoxy)benzoyl)(isobutyl)carbamate (15)**

Following the general procedure for nitro reduction, azide **12** (551 mg, 1.31 mmol) was reduced and purified (2:1 pet:ether) to yield the title compound as a pale tan oil (474 mg, 1.22 mmol, 93 %): δ_H (250 MHz, $CDCl_3$) 7.12 (1H, d, *J* 1.7, Ar), 7.07 (1H, dd, *J* 8.0; 1.8,



Ar), 6.63 (1H, d, *J* 8.1, Ar), 4.12 (2H, t, *J* 6.0, H1), 3.59 (2H, d, *J* 7.3, H5), 3.51 (2H, t, *J* 6.6, H3), 2.16-2.00 (3H, m, H2 and H6), 1.24 (9H, s, H11-13), 0.94 (6H, d, *J* 6.7, H7 and H8). δ_C (62.5 MHz, $CDCl_3$) 173.2 (C4), 154.7 (C9), 145.9 (Ar), 141.7 (Ar), 127.0 (Ar), 123.0 (Ar), 113.0 (Ar), 111.6 (Ar), 82.1 (C10), 65.2 (C1), 53.1 (C5), 48.4 (C3), 28.7 (C6), 28.5 (C2), 27.8 (C11-C13), 20.3 (C7 and C8). IR υ_{max} 3487, 3372, 2962, 2099, 1721, 1633, 1617, 1141. HRMS (ESI) found 414.2101 $C_{19}H_{29}N_5O_4Na^+$ $[M+Na]^+$ requires 414.2112.

tert-Butyl (4-amino-3-(but-3-en-1-yloxy)benzoyl)(isobutyl)carbamate (16)



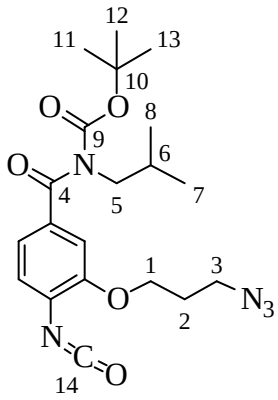
Following the general procedure for nitro reduction, alkene **13** (67 mg, 0.17 mmol, 1.0 eq) was reduced and purified (2:1 petrol:ether) to yield the title compound as a pale pink oil (59 mg, 0.16 mmol, 95 %): δ_H (250 MHz, $CDCl_3$) 7.12 (1H, d, *J* 1.7, Ar), 7.06 (1H, dd, *J* 8.1; 1.7, Ar), 6.62 (1H, d, *J* 8.1, Ar), 5.97-5.79 (1H, m, H3), 5.22-5.06 (2H, m, H4), 4.07 (2H, t, *J* 6.5, H1), 3.59 (2H, d, *J* 7.3, H6), 2.46 (2H, q, *J* 6.6, H2), 2.07 (1H, sept, *J* 6.8, H7), 1.23 (9H, s, H12-14), 0.94 (6H, d, *J* 6.7, H8 and H9). δ_C (62.5 MHz, $CDCl_3$) 173.7 (C5), 155.1 (C10), 145.8 (Ar), 140.8 (Ar), 134.8 (C3), 127.2 (Ar), 123.2 (Ar), 117.6 (C4), 113.2 (Ar), 112.1 (Ar), 84.5 (C11), 67.9 (C1), 53.5 (C6), 34.0 (C2), 28.9 (C7), 28.0 (C12-14), 20.7 (C8 and C9). IR υ_{max} 3490, 3373, 2961, 1720, 1615, 1520, 1240, 1138. HRMS (ESI) found 385.2084 $C_{20}H_{30}N_2O_4Na^+$ $[M+Na]^+$ requires 385.2098.

3.14.5. General Procedure for the Formation of Isocyanates

Sodium bicarbonate (2.0 eq) and bis(trichloromethyl) carbonate (0.6 eq) were added to a vigorously stirred solution of a substituted aniline (1.0 eq) in 1:1 DCM:water (2 mL /100

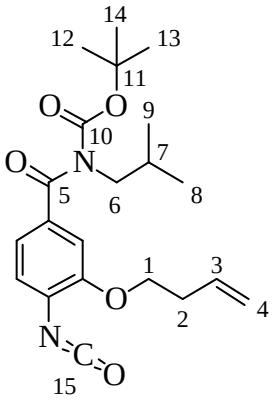
mg) was added. The mixture was stirred for 30 minutes, then diluted with NH_4Cl (10 mL/100 mg) and DCM (10 mL/100 mg). The mixture was extracted with DCM (3 x 10 mL/100mg) and the organic layers combined, dried (Na_2SO_4), filtered, and solvent removed *in vacuo*. The product was immediately used for subsequent reaction without further purification.

***tert*-Butyl (3-(3-azidopropoxy)-4-isocyanatobenzoyl)(isobutyl)carbamate (17)**



Following the general procedure for the formation of isocyanates, reaction with azide **15** (85 mg, 0.22 mmol) afforded the title compound as a yellow oil (102 mg, 0.22 mmol, *quant*): δ_H (250 MHz, $CDCl_3$) 7.26 (1H, d, *J* 1.5, Ar), 7.19 (1H, dd, *J* 8.0; 1.6, Ar), 7.13 (1H, d, *J* 8.1, Ar), 4.32 (2H, t, *J* 5.8, H1), 3.76 (2H, d, *J* 7.4, H5), 3.71 (2H, t, *J* 6.4, H3), 2.33-2.14 (3H, m, H2 and H6), 1.37 (9H, s, H11-13), 1.09 (6H, d, *J* 6.7, H7 and H8). δ_C (62.5 MHz, $CDCl_3$) 172.7 (C4), 154.4 (C9), 152.8 (Ar), 136.2 (Ar), 131.4 (Ar), 126.7 (C14), 123.3 (Ar), 121.0 (Ar), 111.0 (Ar), 83.4 (C10), 66.0 (C1), 53.3 (C5), 48.2 (C3), 28.9 (C6), 28.7 (C2), 27.9 (C11-C13), 20.6 (C7 and C8). IR υ_{max} 2962, 2242, 2098, 1731, 1670, 1332, 1144.

***tert*-Butyl (3-(but-3-en-1-yloxy)-4-isocyanatobenzoyl)(isobutyl)carbamate (18)**



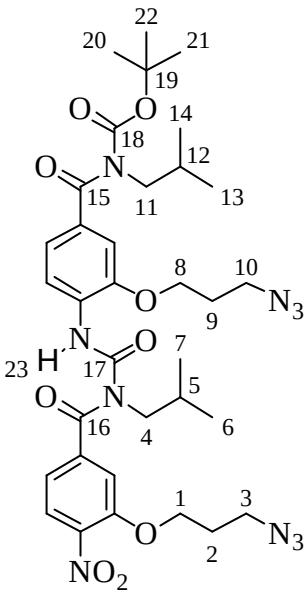
Following the general procedure for the formation of isocyanates, reaction with alkene **16** (71 mg, 0.20 mmol) afforded the title compound as a colourless oil (76 mg, 0.20 mmol, *quant*): δ_H (250 MHz, $CDCl_3$) 7.17 (1H, d, *J* 1.7, Ar), 7.09 (1H, dd, *J* 8.1; 1.7, Ar), 7.03 (1H, d, *J* 6.9, Ar), 6.05-5.86 (1H, m, H3), 5.30-5.15 (2H, m, H4), 4.19 (2H, t, *J* 6.5, H1), 3.68 (2H, d, *J* 7.3, H6), 2.67 (2H, q, *J* 6.6, H2), 2.15 (1H, sept, *J* 6.9, H7), 1.28 (9H, s, H12-14), 1.01 (6H, d, *J* 6.7, H8 and H9). δ_C

(62.5 MHz, $CDCl_3$) 172.9 (C5), 154.5 (Ar), 153.0 (C10), 136.1 (Ar), 134.1 (C3), 131.4 (Ar), 127.2 (C15), 123.2 (Ar), 120.7 (Ar), 117.6 (C4), 111.1 (Ar), 83.3 (C11), 69.0 (C1), 53.3 (C6), 33.7 (C2), 28.7 (C7), 27.9 (C12-14), 20.6 (C8 and C9). IR υ_{max} 2962, 2241, 1731, 1621, 1520, 1332, 1145.

3.14.6. General Procedure for Benzoylurea Formation

Potassium hexamethyldisilazide (0.5 M in toluene, 1.1 eq) was added dropwise to a stirred solution of an amide (1.0 eq) in THF (2 mL/mmol) at -78 ºC. A solution of an isocyanate (1.0 eq) in THF (2 mL/mmol) was added dropwise and the mixture stirred for 2 hours at -78 ºC. AcOH (0.1 mL) was added dropwise, and the solution was allowed to warm to room temperature. The solution was diluted with NH_4Cl (10 mL/100 mg) and DCM (10 mL/100 mg) and extracted with DCM (3 x 10 mL/100 mg). The organic layers were combined, dried (Na_2SO_4), filtered, and purified *via* flash column chromatography.

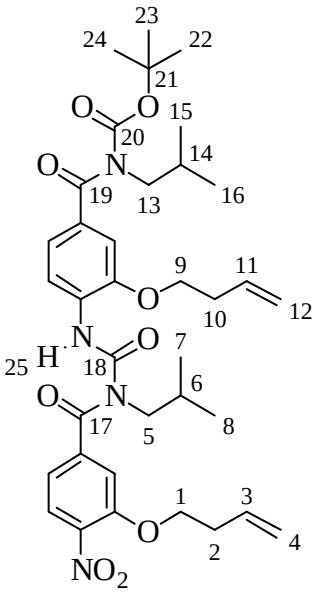
tert-Butyl(3-(3-azidopropoxy)-4-(3-(3-(3-azidopropoxy)-4-nitrobenzoyl)-3-isobutylureido) benzoyl)(isobutyl)carbamate (19)



Following the general procedure for benzoylurea formation, amide **9** (129 mg, 0.4 mmol) was reacted with isocyanate **17** (195 mg, 0.4 mmol) and purified (2:1 petrol:ether) to yield the title compound as a yellow oil (212 mg, 0.29 mmol, 72 %): δ_H (500 MHz, $CDCl_3$) 11.67 (1H, s, H23), 8.37 (1H, d, *J* 8.4, Ar), 7.93 (1H, d, *J* 8.2, Ar), 7.18 (2H, s, Ar), 7.16 (1H, d, *J* 8.8, Ar) 7.12 (1H, d, *J* 8.3, Ar), 4.23 (2H, t, *J* 5.7, H1 or H8), 4.17 (2H, t, *J* 5.7, H1 or H8), 3.69-3.57 (8H, m, H3, H10, H4, and H11), 2.19-2.06 (5H, m, H2, H9, and H12), 1.97 (1H, sep, *J* 6.8, H5), 1.24 (9H, s, H20-22), 0.95 (6H, d, *J* 6.7, H6 and 7 or H13 and 14), 0.81 (6H, d, *J* 6.7, H6 and 7 or H13

and 14). δ_C (125 MHz, $CDCl_3$) 173.5 (C15 or C16), 173.2 (C15 or C16), 154.7 (C18), 152.6 (C17), 151.5 (Ar), 147.7 (Ar), 141.7 (Ar), 141.0 (Ar), 133.3 (Ar), 130.8 (Ar), 126.6 (Ar), 121.7 (Ar), 119.3 (Ar), 119.0 (Ar), 113.8 (Ar), 110.8 (Ar), 83.2 (C19), 66.8 (C1 or C8), 65.8 (C1 or C8), 54.6 (C4 or C11), 53.4 (C4 or C11), 48.5 (C3 or C10), 48.1 (C3 or C10), 29.2 (C5 or C12), 29.1 (C5 or C12), 28.8 (C2 and C9), 28.0 (C20-22), 20.7 (C6-7 or C13-14), 20.3 (C6-7 or C13-14). IR υ_{max} 3221, 2962, 2874, 2098, 1715, 1530, 1331, 1147. HRMS (ESI) found 737.3369 $C_{34}H_{45}N_{10}O_9^+$ $[M+H]^+$ requires 737.3376.

***tert*-Butyl(3-(but-3-en-1-yloxy)-4-(3-(3-(but-3-en-1-yloxy)-4-nitrobenzoyl)-3-isobutylureido) benzoyl)(isobutyl)carbamate (20)**



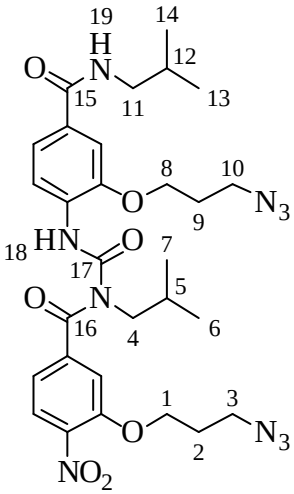
Following the general procedure for benzoylurea formation, amide **10** (493 mg, 1.69 mmol) was reacted with isocyanate **18** (655 mg, 1.69 mmol) and purified (4:1 petrol:ether) to yield the title compound as a yellow oil (726 mg, 1.07 mmol, 63 %): δ_H (500 MHz, $CDCl_3$) 11.47 (1H, s, H25), 8.34 (1H, d, *J* 8.4, Ar), 7.88 (1H, d, *J* 8.2, Ar), 7.21-7.08 (4H, m, Ar), 5.99-5.85 (2H, m, H3 and H11), 5.24-5.06 (4H, m, H4 and H12), 4.18 (2H, t, *J* 6.5, H1 or H9), 4.14 (2H, t, *J* 6.7, H1 or H9), 3.63 (4H, m, H5 and H13), 2.63 (4H, sextet, *J* 6.5, H2 and H10), 2.11 (1H, sept, *J* 6.8, H6 or H14), 1.98 (1H, sept, *J* 6.8, H6 or H14), 1.24 (9H, s, H22-24), 0.96 (6H, d, *J* 6.7, H7 and H8 or H15 and H16), 0.82 (6H, d, *J* 6.6, H7 and H8 or H15 and H16). δ_C (125 MHz, $CDCl_3$) 173.3 (C19 and C17), 154.7 (C18), 154.1 (C20), 152.8 (Ar), 151.7 (Ar), 148.0 (Ar), 141.7 (Ar), 141.3 (Ar), 134.3 (C3 or C11), 133.6 (C3 or C11), 133.3 (Ar), 130.9 (Ar), 126.4 (Ar), 121.5 (Ar), 119.0 (C4 or C12), 118.6 (Ar), 118.0 (C4 or C12), 113.7 (Ar), 111.1 (Ar), 83.2 (C21), 69.6 (C1 or C9), 68.9 (C1 or C9), 54.5 (C5 or C13), 53.4 (C5 or C13), 34.0 (C2

or C10), 33.7 (C2 or C10), 30.1 (C6 or C14), 29.0 (C6 or C14), 28.0 (C22-24), 20.7 (C7 and 8 or C15 and 16), 20.3 (C7 and 8 or C15 and 16).IR υ_{max} 3219, 2961, 2927, 1715, 1532, 1333, 1148. HRMS (ESI) found 703.3314 $C_{36}H_{48}N_4O_9Na^+$ $[M+Na]^+$ requires 703.3314.

3.14.7. General Procedure for N-Boc Deprotection

TFA (1 mL) was added to a stirred solution of an *N*-Boc protected molecule in DCM (1 mL). The mixture was stirred at room temperature and toluene (5 mL) was added. The solvent was removed *in vacuo* and the resulting compound azeotroped three times with chloroform.

3-(3-Azidopropoxy)-*N*-((2-(3-azidopropoxy)-4-(isobutylcarbamoyl)phenyl)carbamoyl)-*N*-isobutyl-4-nitrobenzamide (21)



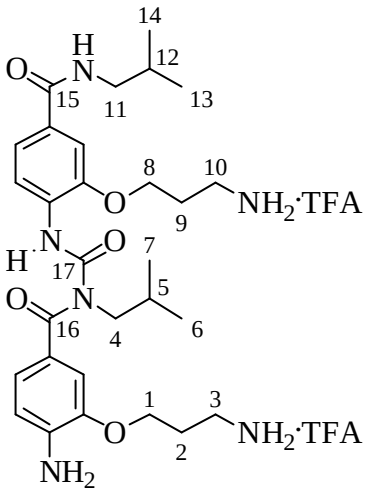
Following the general procedure for *N*-Boc deprotection, benzoylurea **19** (130 mg, 0.18 mmol) was stirred for 5 minutes at 0 ºC and the residue purified *via* flash column chromatography (5:2 ether:petrol) to yield the title compound as an orange oil (94 mg, 0.15 mmol, 84 %): δ_H (250 MHz, $CDCl_3$) 11.63 (1H, s, H18), 8.33 (1H, d, *J* 8.4, Ar), 7.90 (1H, d, *J* 8.3, Ar), 7.46 (1H, d, *J* 1.4, Ar), 7.28 (1H, dd, *J* 8.6; 1.5, Ar), 7.18 (1H, d, *J* 1.1, Ar), 7.11 (1H, dd, *J* 8.3; 1.3, Ar), 6.43 (1H, t, *J* 5.9, H19), 4.21 (2H, t, *J* 5.8, H1 or H8), 4.16 (2H, t, *J* 5.8, H1 or H8), 3.68-3.52 (6H, m, H3, H10, H4), 3.25 (2H, t, *J* 6.4, H11), 2.19-2.02 (4H, m, H2, H9), 2.02-1.78 (2H, m, H5 and H12), 0.95 (6H, d, *J* 6.7, H6 and 7 or H13 and 14), 0.79 (6H, d, *J* 6.7, H6 and 7 or H13 and 14). δ_C (62.5 MHz, $CDCl_3$) 173.4 (C16), 167.3 (C15), 152.6 (C17), 151.6 (Ar), 148.1 (Ar), 141.7 (Ar), 141.0 (Ar), 130.8 (Ar), 130.6 (Ar), 126.5 (Ar), 119.4 (Ar), 119.4 (Ar), 119.3 (Ar), 113.8 (Ar), 110.8 (Ar), 66.9 (C1 or C8), 65.8 (C1 or C8), 54.5 (C4), 48.5 (C3 or C10), 48.1 (C3 or C10), 47.9 (C11), 29.2 (C5 or C12), 29.1 (C5 or C12), 29.0 (C2 or C9), 28.8 (C2 or C9), 20.7 (C6-7 or C13-14), 20.3 (C6-7 or C13-14).

IR υ_{max} 3226, 2960, 2873, 2097, 1710, 1520, 1122. HRMS (ESI) found 661.2817 $C_{29}H_{38}N_{10}O_7Na^+$ $[M+Na]^+$ requires 661.2817.

3.14.8. General Procedure for Catalytic Hydrogenation

Palladium on activated carbon (10 % by mass) was added to a stirred solution of an organic molecule in EtOH. The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen overnight. The solution was filtered through Celite® with DCM (50 mL/mmol) and the solvent removed *in vacuo*.

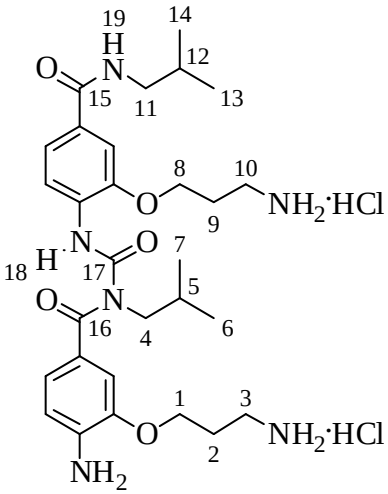
3-(2-(3-(4-Amino-3-(3-ammoniopropoxy)benzoyl)-3-isobutylureido)-5-(isobutylcarbamoyl) phenoxy)propan-1-aminium • 2 TFA (2)



Following the general procedure for catalytic hydrogenation, benzoylurea **21** (41 mg, 0.064 mmol) was hydrogenated overnight to yield the title compound as a brown oil (49 mg). A small amount of the compound was purified *via* semi-preparative HPLC (isocratic 74:26 water:acetonitrile + 0.1 % TFA) to yield a viscous tan oil: δ_H (400 MHz, MeOD) 8.34 (1H, d, *J* 8.4, Ar), 7.54-7.48 (2H, m, Ar), 7.14 (1H, dd, *J* 8.1; 1.8, Ar), 7.10 (1H, d, *J* 1.7, Ar), 6.83 (1H, d, *J* 8.1, Ar), 4.29 (2H, t, *J* 5.6, H1 or H8), 4.16 (2H, t, *J* 5.7, H1 or H8), 3.86 (2H, d, *J* 7.2, H4 or H11), 3.36 (2H, t, *J* 6.5, H3 or H10), 3.28 (2H, t, *J* 6.5, H3 or H10), 3.22 (2H, d, 7.0, H4 or H11), 2.36-2.24 (4H, m, H2 and H9), 1.97 (2H, t, *J* 6.5, H5 and H12), 0.98 (6H, d, *J* 6.6, H6 and H7 or H13 and H14), 0.82 (6H, d, *J* 6.4, H6 and H7 or H13 and H14). δ_C (125 MHz, MeOD) 177.8 (C16), 169.6 (C15), 162.5 (qt, TFA, J_{CF} 35 Hz),154.0 (C17), 148.8 (Ar), 146.5 (Ar), 143.0 (Ar), 132.1 (Ar), 130.8 (Ar), 124.5 (Ar), 124.4 (Ar), 121.3 (Ar), 120.0 (Ar), 114.7

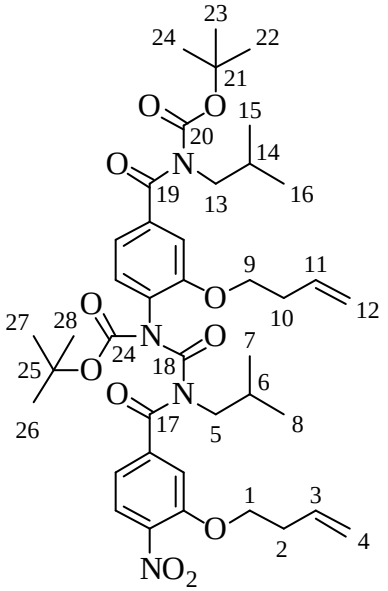
(Ar), 112.6 (Ar), 111.4 (Ar), 66.9 (C1 or C8), 65.8 (C1 or C8), 56.3 (C4 and C11), 38.5 (C3 or C10), 38.4 (C3 or C10), 29.9 (C5 or C12), 29.8 (C5 or C12), 28.4 (C2 and C9), 20.6 (C6-7 or C13-14), 20.2 (C6-7 or C13-14). IR υ_{max} 3369, 2962, 1678, 1628, 1204, 1136. HRMS (ESI) found 557.3442 $C_{29}H_{45}N_6O_5^+$ $[M+H]^+$ requires 557.3446.

3-(2-(3-(4-Amino-3-(3-ammoniopropoxy)benzoyl)-3-isobutylureido)-5-(isobutylcarbamoyl) phenoxy)propan-1-aminium • 2HCl (22)



The TFA salt **2** (34 mg, 0.04 mmol) was dissolved in HCl (0.1 M, 5 mL) stirred for 5 minutes. The solution was frozen at -78 ºC and lyophilized to yield the title compound as a viscous tan oil (29 mg, *quant*). Spectroscopic data matched that of compound **2** except for the absence of the ^{13}C NMR peak at 162.5 ppm and the IR stretch at 1679 cm^{-1}.

***tert*-Butyl(3-(but-3-en-1-yloxy)-4-(3-(3-(but-3-en-1-yloxy)-4-nitrobenzoyl)-1-(*tert*-butoxycarbonyl)-3-isobutylureido)benzoyl)(isobutyl)carbamate (23)**

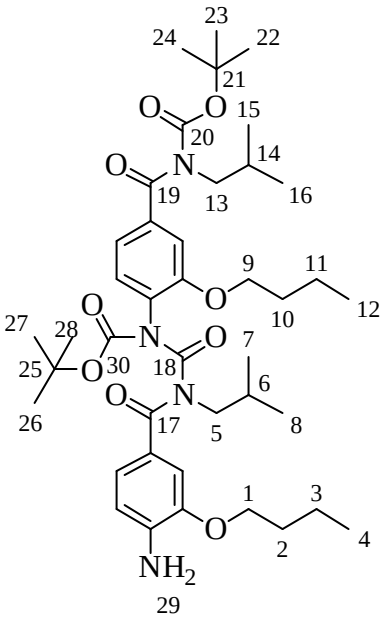


Di-*tert*-butyl dicarbonate (467 mg, 2.14 mmol, 2.0 eq) was added to a stirred solution of benzoylurea **20** (726 mg, 1.07 mmol, 1.0 eq) in THF (25 mL). DMAP (262 mg, 2.14 mmol, 2.0 eq) was added and the mixture stirred for 15 hours at room temperature. The reaction was diluted with NH_4Cl (50 mL) and DCM (50 mL) and extracted with DCM (3 x 50 mL). The organic layers were combined, dried (Na_2SO_4), filtered, concentrated, and purified *via* flash

column chromatography (6:1 petrol:ether) to yield the title compound as a light brown oil (787 mg, 1.01 mmol, 94 %): δ_H (250 MHz, $CDCl_3$) 7.84 (1H, d, *J* 8.3, Ar), 7.41 (1H, d, *J* 1.3, Ar), 7.31 (1H, dd, *J* 8.3; 1.5, Ar), 7.12 (1H, d, *J* 2.0, Ar), 7.01 (1H, dd, *J* 8.0; 1.6, Ar), 6.61 (1H, d, *J* 8.0, Ar), 5.98-5.71 (2H, m, H3 and 11), 5.23-5.04 (4H, m, H4 and H12), 4.15 (2H, t, *J* 6.6, H1 or H9), 4.02 (2H, t, *J* 6.9, H1 or H9), 3.65 (4H, m, H5 and H13), 2.57 (2H, q, *J* 6.6, H2 or H10), 2.45 (2H, q, *J* 6.8, H2 or H10), 2.24 (1H, sept, *J* 6.8, H6 or H14), 2.10 (1H, sept, *J* 6.9, H6 or H14), 1.31 (9H, s, H22-24 or H26-28), 1.19 (9H, s, H22-24 or H26-28), 0.99 (6H, d, *J* 6.7, H7 and H8 or H15 or H16), 0.95 (6H, d, *J* 6.7, H7 and H8 or H15 or H16). δ_C (125 MHz, $CDCl_3$) 172.7 (C17 or C19), 170.5 (C 17 or C19), 156.0 (C18), 154.7 (C20), 154.3 (C24), 152.4 (Ar), 151.6 (Ar), 141.7 (Ar), 141.6 (Ar), 139.8 (Ar), 133.8 (C3 or C11), 133.7 (C3 or C11), 130.5 (Ar), 127.5 (Ar), 125.6 (Ar), 120.2 (Ar), 120.1 (Ar), 118.4 (C4 or C12), 118.1 (C4 or C12), 115.3 (Ar), 112.2 (Ar), 84.5 (C21 or C25), 83.7 (C21 or C25), 69.7 (C1 or C9), 68.6 (C1 or C9), 55.8 (C5 or C13), 53.3 (C5 or C13), 33.8 (C2 or C10), 33.6 (C2 or C10), 30.1 (C6 or C14), 29.0 (C6 or C14), 28.2 (C22-24 or C26-28), 28.0 (C22-24 or C26-28), 20.8 (C7 and C8 or C15 and C16), 20.6 (C7 and C8 or C15 and C16). IR υ_{max} 2961, 2923, 1734, 1671, 1334, 1150. HRMS (ESI) found 803.3838 $C_{41}H_{56}N_4O_{11}Na^+$ $[M+Na]^+$ requires 803.3838.

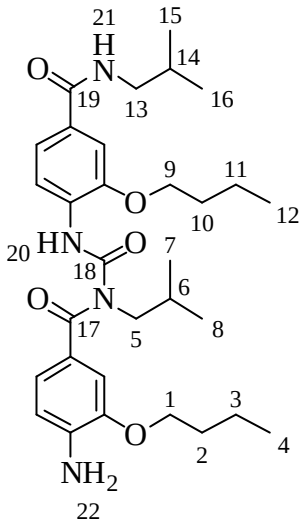
tert-Butyl(4-(3-(4-amino-3-butoxybenzoyl)-1-(_tert_-butoxycarbonyl)-3-isobutylureido)-3-butoxybenzoyl)(isobutyl)carbamate (24)

Following the general procedure for catalytic hydrogenation, benzoylurea **23** (106 mg, 0.14 mmol) was hydrogenated and purified *via* flash column chromatography (5:2 petrol:ether) to yield the title compound (85 mg, 0.11 mmol, 79 %) as a tan oil: δ_H (400 MHz, $CDCl_3$) 8.33 (1H, d, *J* 8.4, Ar), 7.15 (1H, dd, *J* 7.7; 1.7, Ar), 7.07 (1H, d, *J* 1.6, Ar), 6.95 (1H, dd, *J* 8.0; 1.8, Ar), 6.66 (1H, d, *J* 7.8, Ar), 6.59 (1H, d, *J* 8.0, Ar), 4.26 (br s, 2H, H29), 3.98 (2H,



t, *J* 6.4, H1 or H9), 3.93 (2H, t, *J* 6.6, H1 or H9), 3.66 (2H, d, *J* 6.5, H5 or H13), 3.60, (2H, d, *J* 7.3, H5 or H13), 2.21 (1H, sept, *J* 6.8, H6 or H14), 2.09 (1H, sept, *J* 6.8, H6 or H14), 1.75 (2H, quint, *J* 6.5, H2 or H10), 1.68 (2H, quint, *J* 6.8, H2 or H10), 1.54-1.34 (4H, m, H3 and H11), 1.26 (9H, s, H22-24 or H26-28), 1.19 (9H, s, H22-24 or H26-28), 0.99-0.88 (18H, m, H4, H7, H8, H12, H15, and H16). δ_C (100 MHz, $CDCl_3$) 172.6 (C17 or C19), 171.9 (C17 or C19), 156.1 (C18), 154.6 (C20 or C30), 154.6 (C20 or C30), 151.1 (Ar), 145.7 (Ar), 140.8 (Ar), 138.6 (Ar), 130.8 (Ar), 127.6 (Ar), 125.2 (Ar), 123.3 (Ar), 119.2 (Ar), 112.4 (Ar), 112.1 (Ar), 111.3 (Ar), 83.1 (C21 or C25), 82.9 (C21 or C25), 68.4 (C1 or C9), 68.0 (C1 or C9), 55.5 (C5 or C13), 52.8 (C5 or C13), 31.2 (C2 or C10), 31.0 (C2 or C10), 28.2 (C6 or C14), 28.2 (C6 or C14), 27.7 (C22-24 or C26-28), 27.4 (C22-24 or C26-28), 20.4 (C7 and 8 or C15 and 16), 20.2 (C7 and 8 or C15 and 16), 19.3 (C3 or C11), 19.0 (C3 or C11), 13.8 (C4 or C12), 13.7 (C4 or C12). IR υ_{max} 3374, 2960, 2934, 1733, 1699, 1359. HRMS (ESI) found 777.4425 $C_{41}H_{62}N_4O_9Na^+$ $[M+Na]^+$ requires 777.4409.

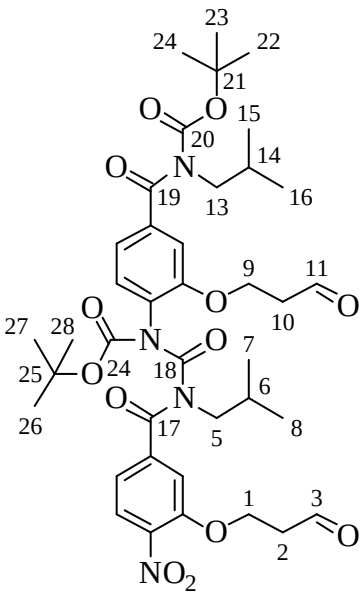
4-Amino-3-butoxy-*N*-((2-butoxy-4-(isobutylcarbamoyl)phenyl)carbamoyl)-*N*-isobutylbenzamide (4)



Following the general procedure for *N*-Boc deprotection, benzoylurea **24** (73 mg, 0.097 mmol was stirred for 5 minutes at 0 ºC and the residue purified *via* flash column chromatography (2:1 ether:petrol) to yield the title compound as a thick tan oil (41 mg, 0.074 mmol, 76 %): δ_H (400 MHz, $CDCl_3$) 11.16 (1H, s,

H20), 8.36 (1H, d, *J* 8.4, Ar), 7.45 (1H, d, *J* 1.7, Ar), 7.22 (1H, dd, *J* 8.4; 1.8, Ar), 7.03 (1H, dd, *J* 8.0; 1.7, Ar), 7.00 (1H, d, *J* 1.6, Ar), 6.66 (1H, d, *J* 8.0, Ar), 6.29 (1H, t, *J* 5.7, H21), 4.22 (2H, s br, H22), 4.07 (2H, t, *J* 6.7, H1 or H9), 4.01 (2H, t, *J* 6.5, H1 or H9), 3.80 (2H, d, *J* 7.2, H5), 3.27 (2H, t, *J* 6.4, H13), 1.99 (1H, sept, *J* 6.8, H6), 1.95-1.75 (5H, m, H2, H10, H14), 1.50 (4H, dq, *J* 7.5; 2.3, H3 and H11), 1.01-0.93 (12H, m, H4, H12, H15-H16), 0.80 (6H, d, *J* 6.7, H7-H8). δ_C (125 MHz, $CDCl_3$) 175.6 (C17 or C19), 167.6 (C17 or C19), 153.3 (C18), 148.5 (Ar), 146.2 (Ar), 140.5 (Ar), 131.6 (Ar), 129.9 (Ar), 124.9 (Ar), 122.8 (Ar), 119.0 (Ar), 118.8 (Ar), 113.6 (Ar), 111.8 (Ar), 110.9 (Ar), 69.1 (C1 or C9), 68.7 (C1 or C9), 55.4 (C5), 47.8 (C13), 31.7 (C2 or C10), 31.5 (C2 or C10), 29.1 (C14), 29.0 (C6), 20.6 (C15-16), 20.4 (C7-8), 19.7 (C3 or C11), 19.5 (C3 or C11), 14.3 (C4 or C12), 14.2 (C4 or C12). IR υ_{max} 3487, 3351, 2959, 2935, 1702, 1520, 1359. HRMS (ESI) found 577.3352 $C_{31}H_{46}N_4O_5Na^+$ $[M+Na]^+$ requires 577.3360.

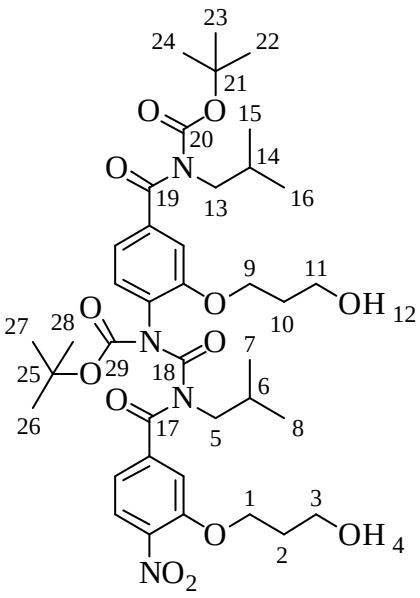
***tert*-Butyl(4-(1-(*tert*-butoxycarbonyl)-3-isobutyl-3-(4-nitro-3-(3-oxopropoxy)benzoyl) ureido)-3-(3-oxopropoxy)benzoyl)(isobutyl)carbamate (25)**



Benzoylurea **20** (92 mg, 0.12 mmol, 1 eq) was dissolved in DCM (2.5 mL) and methanol (0.5 mL) and cooled to -78 ºC. The solution was flushed with O_2 for 5 minutes, then O_3 was flushed into the solution until it turned a pale blue colour. The solution was flushed for 4 minutes with O_2, then DMS (37 mg, 0.59 mmol, 5 eq) was added and the solution stirred at room temperature overnight. The solution was diluted with pH 7 buffer (20 mL) and extracted with DCM (3 x 20 mL). The organic layers were combined, dried (Na_2SO_4), filtered, and solvent removed *in vacuo* to yield the title compound as a yellow oil (78 mg,

0.10 mmol, 85 %): δ_H (500 MHz, $CDCl_3$) 9.83 (1H, s, H3 or H11), 9.74 (1H, s, H3 or H11), 7.84 (1H, d, *J* 8.3, Ar), 7.41 (1H, s, Ar), 7.31 (1H, d, *J* 8.3, Ar), 7.15 (1H, s, Ar), 7.04 (1H, d, *J* 7.9, Ar), 6.65 (1H, d, *J* 7.6, Ar), 4.42 (2H, t, *J* 5.9, H1 or H9), 4.32 (2H, t, *J* 6.1, H1 or H9), 3.66-3.56 (4H, m, H5 and H13), 2.98 (2H, t, *J* 5.8, H2 or H10), 2.80 (2H, t, *J* 6.0, H2 or H10), 2.19 (1H, sept, *J* 6.8, H6 or H14), 2.07 (1H, sept, *J* 6.8, H6 or H14), 1.30 (9H, s, H22-24 or H26-28), 1.18 (9H, s, H22-24 or H26-28), 0.96 (6H, d, *J* 6.6, H7 and H8 or H15 or H16), 0.92 (6H, d, *J* 6.7, H7 and H8 or H15 or H16). δ_C (125 MHz, $CDCl_3$) 199.4 (C3 or C11), 199.0 (C3 or C11), 172.5 (C17 or C19), 170.4 (C 17 or C19), 155.9 (C18), 154.3 (C20 or 24), 154.2 (C20 or 24), 152.0 (Ar), 151.5 (Ar), 141.7 (Ar), 141.6 (Ar), 139.9 (Ar), 130.4 (Ar), 127.8 (Ar), 125.7 (Ar), 120.8 (Ar), 120.6 (Ar), 115.4 (Ar), 112.4 (Ar), 84.8 (C21 or C25), 83.7 (C21 or C25), 64.0 (C1 or C9), 63.0 (C1 or C9), 55.8 (C5 or C13), 53.3 (C5 or C13), 43.2 (C2 or C10), 43.0 (C2 or C10), 28.8 (C6 or C14), 28.6 (C6 or C14), 28.1 (C22-24 or C26-28), 27.9 (C22-24 or C26-28), 20.7 (C7 and 8 or C15 and 16), 20.6 (C7 and 8 or C15 and 16). IR υ_{max} 2963, 2873, 1733, 1671, 1335, 1151.

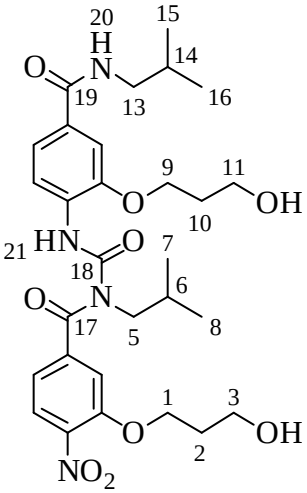
***tert*-Butyl(4-(1-(*tert*-butoxycarbonyl)-3-(3-(3-hydroxypropoxy)-4-nitrobenzoyl)-3-isobutylureido)-3-(3-hydroxypropoxy)benzoyl)(isobutyl)carbamate (26)**



Bis-aldehyde **25** (37 mg, 0.047 mmol, 1.0 eq) was dissolved in MeOH (2 mL) and sodium borohydride (3 mg, 0.075 mmol, 1.6 eq) was added. The solution was stirred for 40 minutes, after which time DCM (10 mL) and NH_4Cl (10 mL) were added. The solution was extracted with DCM (3 x 10 mL), dried (Na_2SO_4), and purified *via* flash column chromatography (8:1 ether:petrol) to yield the product as a colourless oil (13

mg, 0.016 mmol, 35 %): δ_H (500 MHz, $CDCl_3$) 7.89 (1H, d, *J* 8.3, Ar), 7.48 (1H, s, Ar), 7.33 (1H, d, *J* 8.3, Ar), 7.15 (1H, s, Ar), 7.02 (1H, dd, *J* 8.0; 1.2, Ar), 6.97 (1H, s, Ar), 4.30 (2H, t, *J* 5.8, H1 or H9), 4.14 (2H, t, *J* 6.2, H1 or H9), 3.86 (2H, t, *J* 5.6, H3 or H11), 3.70 (2H, t, *J* 5.8, H3 or H11), 3.65 (2H, s br, H5 or H13), 3.61 (2H, d, *J* 7.4, H5 or H13), 2.23 (1H, sept, *J* 6.9, H6 or H14), 2.15-2.03 (3H, m, H6 or H14 and H2 or H10), 1.94 (2H, quint, *J* 6.0, H2 or H10), 1.35 (9H, s, H22-24 or H26-28), 1.20 (9H, s, H22-24 or H26-28), 0.98 (6H, d, *J* 6.7, H7 and H8 or H15 or H16), 0.93 (6H, d, *J* 6.7, H7 and H8 or H15 or H16). δ_C (125 MHz, $CDCl_3$) 172.7 (C17 or C19), 170.7 (C 17 or C19), 154.7 (C18), 154.3 (C20 or 29), 154.5 (C20 or 29), 151.9 (Ar), 141.8 (Ar), 139.9 (Ar), 136.2 (Ar), 130.3 (Ar), 127.7 (Ar), 126.0 (Ar), 125.9 (Ar), 120.4 (Ar), 120.0 (Ar), 115.0 (Ar), 112.3 (Ar), 84.7 (C21 or C25), 83.8 (C21 or C25), 67.9 (C1 or C9), 66.5 (C1 or C9), 59.9 (C3 or C11), 59.6 (C3 or C11), 55.8 (C5 or C13), 53.3 (C5 or C13), 32.1 (C2 or C10), 31.9 (C2 or C10), 28.8 (C6 or C14), 28.6 (C6 or C14), 28.2 (C22-24 or C26-28), 27.9 (C22-24 or C26-28), 20.7 (C7-8 or C15-16), 20.6 (C7-8 or C15-16). IR υ_{max} 3452, 2958, 2873, 1735, 1669, 1335, 1149. HRMS (ESI) found 811.3747 $C_{39}H_{56}N_4O_{13}Na^+$ $[M+Na]^+$ requires 811.3736.

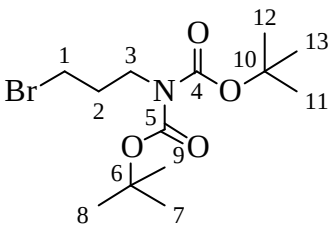
3-(3-Hydroxypropoxy)-*N*-((2-(3-hydroxypropoxy)-4-(isobutylcarbamoyl)phenyl)carbamoyl)-*N*-isobutyl-4-nitrobenzamide (3)



Following the general procedure for *N*-Boc deprotection, diol **26** (13 mg, 0.016 mmol) was stirred for 45 minutes and purified *via* flash column chromatography (100:3 DCM:MeOH) to yield the title compound as a yellow oil (3 mg, 5 mmol, 31 %): δ_H (500 MHz, $CDCl_3$) 11.48 (1H, s, H21), 8.35 (1H, d, *J* 8.4, Ar), 7.93 (1H, d, *J* 8.3, Ar), 7.51 (1H, d, *J* 1.7, Ar), 7.25 (2H, s, Ar), 7.12 (1H, dd, *J* 8.2; 1.5, Ar), 6.21 (1H, t, *J* 5.8, H20), 4.33 (2H, t, *J* 5.9,

H1 or H9), 4.28 (2H, t, *J* 5.9, H1 or H9), 3.93-3.87 (4H, m, H3 and H11), 3.66 (2H, d, *J* 7.1, H5), 3.29 (2H, t, *J* 6.4, H13), 2.17-2.08 (4H, m, H2 and H10), 1.98 (1H, quint, *J* 6.8, H6), 1.91 (1H, quint, *J* 6.8, H14), 0.99 (6H, d, *J* 6.7, H15-16), 0.83 (6H, dd, *J* 6.6; 1.8, H7-8). δ_C (125 MHz, $CDCl_3$) 173.0 (C17), 166.9 (C19), 152.1 (C18), 151.2 (Ar), 148.0 (Ar), 141.4 (Ar), 140.4 (Ar), 130.4 (Ar), 130.3 (Ar), 126.2 (Ar), 118.9 (Ar), 118.7 (Ar), 118.5 (Ar), 113.3 (Ar), 111.7 (Ar), 68.6 (C1 or C9), 66.2 (C1 or C9), 59.8 (C3 or C11), 59.5 (C3 or C11), 54.1 (C5), 47.4 (C13), 31.9 (C2 or C10), 31.5 (C2 or C10), 28.7 (C6 or C14), 28.6 (C6 or C14), 20.1 (C7-8 or C15-16), 19.9 (C7-8 or C15-16). IR υ_{max} 3343, 2960, 2873, 1711, 1714, 1607, 1588 1525. HRMS (ESI) found 611.2690 $C_{29}H_{40}N_4O_9Na^+$ $[M+Na]^+$ requires 611.2687.

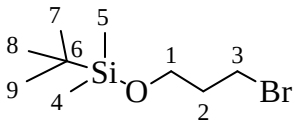
3-Bromopropylamine-*N*,*N*-dicarboxylate (27)



According to the procedure of Dutton *et al*,[28] di-*tert*-butyl iminodicarboxylate (969 mg, 4.46 mmol, 1.0 eq) was dissolved in DMF (20 mL) and THF (20 mL). Sodium hydride (137 mg, 4.57 mmol as 60 % dispersion in oil, 1.0 eq) was added and the mixture heated to 65 ºC. To the solution was added 1,3-dibromo propane (2 mL, 19.7 mmol, 4.0 eq) and the mixture heated for 3 hours. After cooling to room temperature, water (30 mL) and ether (30 mL) were added. The solution was washed with water (2 x 30 mL) and organic layer separated, washed with brine (20 mL), dried ($MgSO_4$), concentrated, and purified *via* flash column chromatography (20:1 petrol:ether) to yield the title compound (0.925 g, 2.73 mmol, 61 %) as a colourless oil: δ_H (400 MHz, $CDCl_3$) 3.70 (2H, t, *J* 7.1, H3), 3.38 (2H, t, *J* 6.6, H1), 2.12 (2H, quint, *J* 6.8, H2), 1.49 (18H, s, H7-9 and H11-13). δ_C (100 MHz, $CDCl_3$) 152.4 (C4 and C5), 82.5 (C6 and C10), 45.2 (C3), 32.1 (C1), 30.5 (C2),

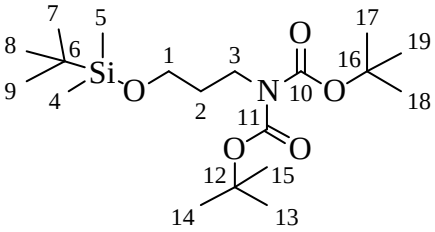
28.1 (C7-9 and C11-13). IR υ_{max} 2980, 1790, 1697, 1394, 1367, 1134. Spectroscopic measurements agree with literature values.[28]

(3-Bromopropoxy)(*tert*-butyl)dimethylsilane (28)



According to the procedure of Trapella *et al*,[29] 3-bromo-propanol (1 mL, 11.44 mmol, 1.0 eq) was added to a stirred solution of *tert*-butyl dimethychloro silane (2.24 g, 14.9 mmol, 1.3 eq) and imidazole (2.34 g, 34.3 mmol, 3.0 eq) in THF (20 mL). The mixture was stirred for 24 hours before NH_4Cl (40 mL) and ether (40 mL) were added. The mixture was extracted with ether (2 x 40 mL) and the organic layers combined, dried ($MgSO_4$), filtered, and purified *via* flash column chromatography (9:1 petrol:ethyl acetate) to yield the title compound (1.26 g, 5.0 mmol, 44 %) as a colourless oil: δ_H (400 MHz, $CDCl_3$) 3.74 (2H, t, *J* 5.7, H1), 3.53 (2H, t, *J* 6.5, H3), 2.04 (2H, quint, *J* 6.0, H2), 0.91 (9H, s, H7-9), 0.08 (6H, s, H4-5). δ_C (100 MHz, $CDCl_3$) 60.4 (C1), 35.5 (C3), 30.7 (C2), 25.9 (C7-9), 18.3 (C6), -5.4 (C4-5). IR υ_{max} 2930, 1257, 1103, 836, 777. Spectroscopic measurements agreed with literature values.[29]

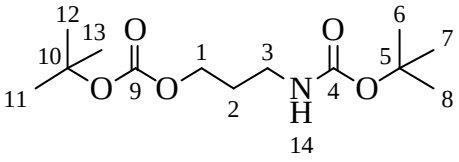
3-(Bis(*tert*-butoxycarbonyl)amino)propoxy)(*tert*-butyl)dimethylsilane (29)



Di-*tert*-butyl iminodicarboxylate (215 mg, 0.99 mmol, 1.25 eq) was dissolved in DMF (5 mL) and THF (5 mL). Sodium hydride (40 mg, 0.99 mmol as 60 % dispersion in oil, 1.25 eq) was added and the mixture heated to 65 ºC. Protected alcohol **28** (200 mg, 0.79 mmol, 1.0 eq) was added to the solution and the mixture heated for 4 hours. After cooling to room temperature, water (20 mL) and ether (20 mL) were added. The solution was washed with water (2 x 30 mL) and organic layer separated, washed with brine (20 mL), dried ($MgSO_4$), concentrated, and purified *via*

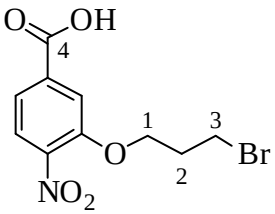
flash column chromatography (12:1 petrol:ether) to yield the title compound (183 mg, 0.47 mmol, 59 %) as a colourless oil: δ_H (400 MHz, $CDCl_3$) 3.67-3.61 (4H, m, H1 and H3), 1.80 (2H, quint, *J* 7.2, H2), 1.51 (18H, s, H13-15 and H17-19), 0.90 (9H, s, H7-9), 0.05 (6H, s, H4-5). δ_C (100 MHz, $CDCl_3$) 152.5 (C10 and C11), 82.0 (C12 and C16), 61.1 (C1), 44.1 (C3), 32.4 (C2), 28.1 (C13-15 and C17-19), 25.9 (C7-9), 18.3 (C6), -5.4 (C4-5). IR υ_{max} 2931, 1748, 1699, 1393, 1160. HRMS (ESI) found 412.2495 $C_{19}H_{39}NO_5SiNa^+$ $[M+Na]^+$ requires 412.2490.

tert-Butyl (3-((_tert_-butoxycarbonyl)oxy)propyl)carbamate (30)



Protected alchohol **29** (100 mg, 0. 26 mmol, 1.0 equiv) was added to a solution of tetra-*n*-butylammonium fluoride (1.0 M solution in THF, 0.28 mL, 0.28 mmol, 1.1 equiv) in THF (1 mL). The solution was stirred overnight, after which time it was concentrated and purified *via* flash column chromatograpy (4:1 petrol: ethyl acetate) to yield the title compound (91 mg, 0.35 mmol, 25 %) as a yellow oil: δ_H (400 MHz, $CDCl_3$) 4.73 (1H, s br, H14), 4.12 (2H, t, *J* 6.2, H1), 3.20 (2H, d, *J* 5.9, H3), 1.84 (2H, quint, *J* 6.4, H2), 1.48 (9H, s, H6-8 or H11-13), 1.43 (9H, s, H6-8 or H11-13). δ_C (100 MHz, $CDCl_3$) 155.9 (C4 or C9), 153.6 (C4 or C9), 82.1 (C5 or C10), 79.2 (C5 or C10), 64.4 (C1), 37.2 (C3), 29.7 (C2), 28.3 (C6-8 or C11-13), 27.7 (C6-8 or C11-13). IR υ_{max} 3371, 2943, 2853, 1743, 1519, 1368, 1280. HRMS (ESI) found 298.1631 $C_{13}H_{25}NO_5Na^+$ $[M+Na]^+$ requires 298.1625.

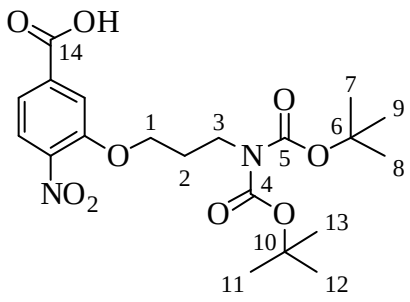
3-(3-Bromopropoxy)-4-nitrobenzoic acid (32)



Adapting the procedure of Boger *et al*,[27] 3-fluoro-4-nitrobenzoic acid **6** (1.00 g, 5.4 mmol, 1.0 eq) was dissolved in THF (40 mL) and

cooled to 0 ºC. Sodium hydride (670 mg, 16.7 mmol, 3.1 eq) was added in portions and the mixture stirred for 5 minutes before 3-bromo-1-propanol (1.5 g, 10.8 mmol, 2 eq) was added over 3 hours by syringe pump. The reaction was then warmed to room temperature and stirred overnight, before being diluted with EtOAc (100 mL). HCl (1M, 60 mL) was added, and the mixture was washed with HCl (0.5 M, 20 mL x 3). The organic layer was collected, dried ($MgSO_4$), filtered, and concentrated, and the product purified *via* flash column chromatography (3:2 Pet:EtOAc + 1 % AcOH) to yield the title compound as a pale yellow solid (1.44 g, 4.7 mmol, 87 %): MP 124-127. δ_H (400 MHz, MeOD) 7.87-7.82 (2H, m, Ar), 7.72 (1H, dd, *J* 8.3; 1.6, Ar), 4.32 (2H, t, *J* 5.7, H1), 3.66 (2H, t, *J* 6.4, H3), 2.35 (2H, quint, *J* 6.1, H2). δ_C (100 MHz, MeOD) 166.5 (C4), 151.6 (Ar), 136.0 (Ar), 132.3 (Ar), 125.0 (Ar), 121.8 (Ar), 115.7 (Ar), 67.4 (C1), 32.2 (C2), 29.2 (C3). IR υ_{max} 2960, 2886, 2605, 1726, 1688, 1525, 1304, 1251, 743. HRMS (ESI) found 301.9677, 303.9659 $C_{10}H_9BrNO_5^-$ $[M-H]^-$ requires 301.9670, 303.9650.

3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy-4-nitrobenzoic acid (31)



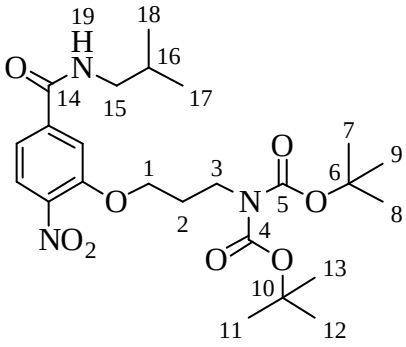
Di-*tert*-butyl iminodicarboxylate (633 mg, 2.92 mmol, 1.1 eq) was added to a solution of sodium hydride (72 mg, 3 mmol, 1.1 eq) in DMF (10 mL) and the mixture heated to 65 ºC. Alkyl bromide acid **32** was dissolved in THF (10 mL) and sodium hydride (72 mg, 3 mmol as a 60 % dispersion in oil, 1.1 eq) added. This solution was stirred for 5 minutes then added *via* syringe to the heated solution containing the di-*tert*-butyl iminodicarboxylate. The mixture was stirred overnight at 65 ºC before being allowed to cool to room temperature. The mixture was diluted with EtOAc, washed with HCl (0.2 M, 10mL x 3), and the organic layer dried (Na_2SO_4), filtered, concentrated, and purified *via* flash column chromatography

(3:2 EtOAc:petrol + 1 % AcOH) to yield the title compound as a viscous yellow oil (1.08 g, 2.45 mmol, 92 %): δ_H (400 MHz, MeOD) 7.87-7.80 (2H, m, Ar), 7.71 (1H, t, *J* 7.7, Ar), 4.25 (2H, t, *J* 5.8, H1), 3.81 (2H, t, *J* 6.9, H3), 2.11 (2H, quint, *J* 6.2, H2), 1.48 (18H, s, H7-9, H11-13). δ_C (100 MHz, MeOD) 166.6 (C14), 153.0 (Ar), 151.7 (C4 and C5), 143.0 (Ar), 135.8 (Ar), 125.0 (Ar), 121.7 (Ar), 115.5 (Ar), 82.9 (C6 and C10), 67.8 (C1), 43.9 (C3), 28.7 (C2), 27.3 (C7-9, C11-13). IR υ_{max} 3300, 2980, 1785, 1704, 1493, 1367, 1254, 1099. HRMS (ESI) found 439.1703 $C_{20}H_{27}N_2O_9^-$ $[M-H]^-$ requires 439.1722.

3.14.9. General Procedure II for Amidation

DIPEA (2.5 eq) was added dropwise to a solution of a carboxylic acid (1.0 eq) in DCM. Mukaiyama's reagent (2-chloro-1-methylpyridinium iodide, 1.2 eq) was added to the mixture and the reaction heated to 40 ºC for 30 minutes. Amine (1.1 eq) was then added and the reaction stirred for 20 hours at 40 ºC. The solution was diluted with NH_4Cl (10 mL/100 mg) and DCM (10 mL/100 mg) and extracted with DCM (3 x 10 mL/100 mg). The organic layers were combined, dried (Na_2SO_4), filtered, and purified *via* flash column chromatography.

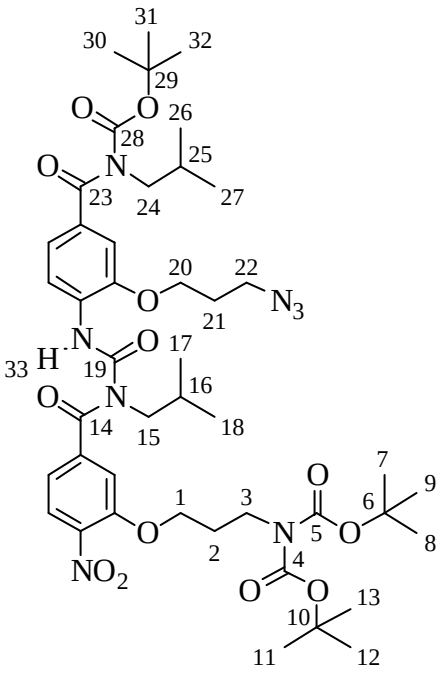
3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy- *N*-isobutyl-4-nitrobenzamide (33)



Following the general procedure II for amidation, di-Boc acid **31** (960 mg, 2.18 mmol) was coupled with isobutyl amine (207 mg, 2.83 mmol, 1.3 eq) and purified (3:2 petrol:ether) to yield the title compound as a pale oil (729 mg, 1.47 mmol, 67 %): δ_H (400 MHz, $CDCl_3$) 7.82 (1H, d, *J* 8.3, Ar), 7.56 (1H, d, *J* 1.4, Ar), 7.29 (1H, dd, *J* 8.3; 1.5, Ar), 6.44 (1H, t, *J* 5.3, H19), 4.20 (2H, t, *J* 6.1, H1), 3.79 (2H, t, *J* 7.1, H3), 3.28 (2H, t, *J* 6.7, H15), 2.11 (2H, quint, *J* 6.3, H2), 1.91 (1H, sept, *J* 6.7, H16), 1.47 (18H, s, H7-9 and H11-13), 0.97 (6H, d, *J* 6.7, H17 and H18). δ_C (100 MHz, $CDCl_3$) 165.5 (C14), 152.5 (C4 and C5), 152.3 (Ar), 141.3 (Ar),

140.0 (Ar), 125.6 (Ar), 117.8 (Ar), 114.0 (Ar), 82.5 (C6 and C10), 67.8 (C1), 47.6 (C15), 43.6 (C3), 28.6 (C2), 28.5 (C16), 28.0 (C7-9, C11-13), 20.1 (C17 and C18). IR υ_{max} 3337, 3084, 2963, 1772, 1645, 1528, 1366, 1110. HRMS (ESI) found 518.2473 $C_{24}H_{37}N_3O_8Na^+$ $[M+Na]^+$ requires 518.2473.

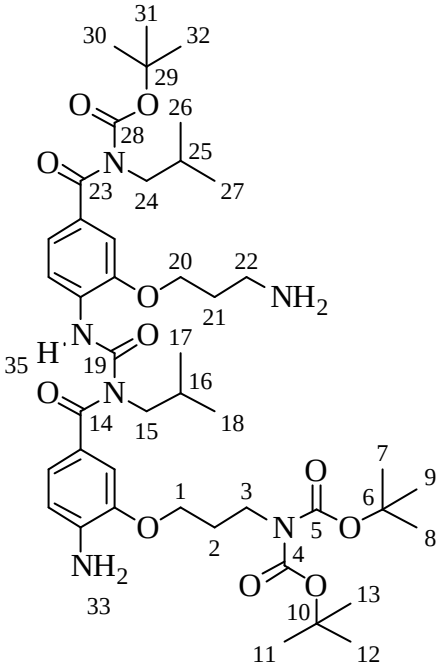
***tert*-Butyl(3-(3-azidopropoxy)-4-(3-(3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)-4-nitrobenzoyl)-3-isobutylureido)benzoyl)(isobutyl)carbamate (34)**



Following the general procedure for benzoylurea formation, amide **33** (521 mg, 1.05 mmol) was reacted with isocyanate **17** (429 mg, 1.03 mmol) and purified (1:1 petrol:ether) to yield the title compound as a viscous orange oil (835 mg, 0.89 mmol, 86 %): δ_H (500 MHz, $CDCl_3$) 11.61 (1H, s, H33), 8.37 (1H, d, *J* 8.3, Ar), 7.91 (1H, d, *J* 8.3, Ar), 7.21-7.14 (3H, m, Ar), 7.11 (1H, dd, *J* 8.2; 1.4, Ar), 4.18 (4H, t, *J* 5.8, H1 and H20), 3.82 (2H, t, *J* 6.9, H3), 3.69-3.62 (6H, m, H15, H22, H24), 2.20-2.06 (5H, m, H2, H21, and H16 or H25), 1.98 (1H, sept, *J* 6.8, H16 or H25), 1.49 (18H, s, H7-9 and H11-13), 1.26 (9H, s, H30-32), 0.97 (6H, d, *J* 6.8, H17 and H18 or H26 and H27), 0.82 (6H, d, *J* 6.7, H17 and H18 or H26 and H27). δ_C (125 MHz, $CDCl_3$) 173.2 (C14 or C23), 172.7 (C14 or C23), 154.2 (C19), 152.6 (C4 and C5 or C28), 152.5 (C4 and C5 or C28), 152.4 (Ar), 151.1 (Ar), 147.3 (Ar), 141.1 (Ar), 140.7 (Ar), 132.8 (Ar), 130.4 (Ar), 125.9 (Ar), 121.2 (Ar), 118.6 (Ar), 113.3 (Ar), 110.4 (Ar), 82.7 (C29), 82.5 (C6 and C10), 68.1 (C1 or C20), 65.4 (C1 or C20), 54.1 (C15 or C24), 53.0 (C15 or C24), 48.0 (C22), 43.5 (C3), 28.7 (C16 or C25), 28.6 (C2 and C21), 28.3 (C16 or C25), 28.0 (C7-9 and C11-13), 27.5 (C30-32), 20.2 (C17 and C18 or

C26 and C27), 19.9 (C17 and C18 or C26 and C27). IR υ_{max} 3223, 2964, 2099, 1716, 1531, 1367, 1333, 1148, 1039. HRMS (ESI) found 935.4472 $C_{44}H_{64}N_8O_{13}Na^+$ $[M+Na]^+$ requires 935.4485.

***tert*-Butyl(4-(3-(4-amino-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)-3-isobutylureido)-3-(3-aminopropoxy)benzoyl)(isobutyl)carbamate (35)**



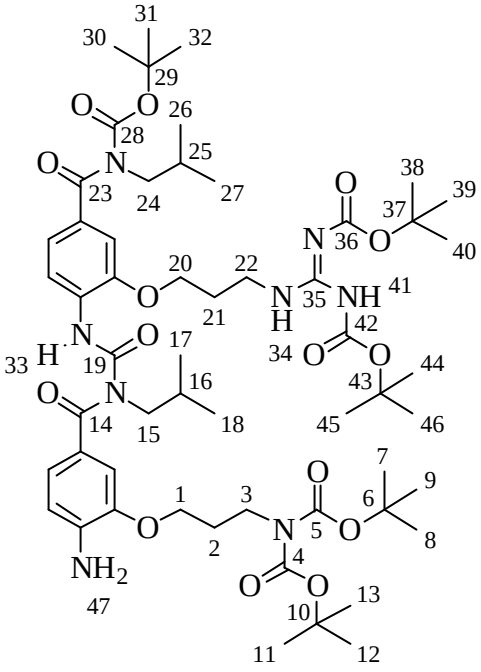
Following the general procedure for catalytic hydrogenation, benzoylurea **34** (181 mg, 0.19 mmol) was hydrogenated to yield the title compound as a tan oil (162 mg, 0.19 mmol, *quant*), which was used for subsequent reactions without further purification: δ_H (500 MHz, $CDCl_3$) 11.43 (1H, s, H35), 8.35 (1H, d, *J* 8.4, Ar), 7.18-7.10 (2H, m, Ar), 7.01 (1H, d, *J* 8.0, Ar), 6.95 (1H, s, Ar), 6.64 (1H, d, *J* 8.0, Ar), 4.31 (1H, s br, H33), 4.14 (2H, t *J* 6.0, H20), 4.06 (2H, t, *J* 5.7, H1), 3.83 (2H, t, *J* 6.5, H3), 3.78 (2H, d, *J* 8.0, H15 or 24), 3.61 (2H, d, *J* 7.3, H15 or 24), 2.95 (2H, t, *J* 6.8, H22), 2.15-1.92 (6H, m, H2, H16, H21, H25), 1.47 (18H, s H7-9 and H11-13), 1.21 (9H, s, H30-32), 0.95 (6H, d, *J* 6.7, H17 and H18 or H26 and H27), 0.77 (6H, d, *J* 6.6, H17 and H18 or H26 and H27). δ_C (125 MHz, $CDCl_3$) 176.0 (C14 or C23), 173.5 (C14 or C23), 154.8 (C19), 153.1 (C4 and C5 or C28), 153.2 (C4 and C5 or C28), 147.9 (Ar), 146.0 (Ar), 140.7 (Ar), 132.5 (Ar), 131.8 (Ar), 124.7 (Ar), 123.0 (Ar), 121.6 (Ar), 118.7 (Ar), 113.7 (Ar), 111.5 (Ar), 111.0 (Ar), 83.1 (C29 or C6 and C10), 82.5 (C29 or C6 and C10), 67.3 (C20), 66.8 (C1), 55.6 (C15 or C24), 53.4 (C15 or C24), 44.3 (C3), 39.5 (C22), 33.3 (C21), 30.8 (C2), 29.0 (C16 or 25), 28.8 (C16 or C25), 28.5 (C7-9 and C11-13), 28.0 (C30-32), 20.7 (C17 and C18 or C26 and C27), 20.4

(C17 and C18 or C26 and C27). IR υ_{max} 3371, 2962, 1729, 1529, 1368, 1233, 1125. HRMS (ESI) found 857.5023 $C_{44}H_{69}N_6O_{11}^+$ $[M+H]^+$ requires 857.5019.

3.14.10. General Procedure for Protected Guanidino Group Formation

N,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine (1.2 eq) was added to a solution of a primary amine (1 eq) in DCM. The reaction was stirred for 40 hours before water (10 mL/100 mg) and DCM (10 mL/100 mg) were added. The mixture was extracted with DCM (2 x 10 mL/100 mg), dried ($MgSO_4$), filtered, and purified *via* flash column chromatography.

tert-Butyl(4-(3-(4-amino-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)-3-isobutylureido)-3-(3-((2,2,10,10-tetramethyl-4,8-dioxo-3,9-dioxa-5,7-diazaundecan-6-yl)amino)propoxy)benzoyl)(isobutyl)carbamate (36)



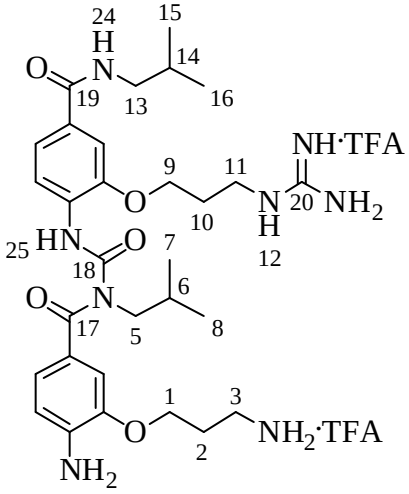
Method 1: *N*,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (65 mg, 0.21 mmol, 1.1 equiv) was added to a stirred solution of amine **35** (162 mg, 0.19 mmol, 1.0 eq) in DCM (3 mL). The reaction was stirred overnight, before water (10 mL) and DCM (10 mL) were added. The mixture was washed with DCM (3 x 10 mL), dried ($MgSO_4$) and purified *via* flash column chromatography (3:2 petrol:ether) to yield the title compound as a viscous, cloudy oil (90 mg, 0.082 mmol, 44 %).

Method 2: *N*,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (171 mg, 0.55 mmol, 1.2 eq) and palladium on activated carbon (210 mg, 5 % Pd by mass) were added to a stirred solution of benzoylurea **35** (420 mg, 0.46 mmol, 1.0 eq) in EtOAc (20 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 69

hours. The solution was filtered through Celite® with EtOAc (200 mL), concentrated, and purified *via* flash column chromatography (1:1 ether:petrol) to yield the title compound as a white foam (442 mg, 0.40 mmol, 87 %).

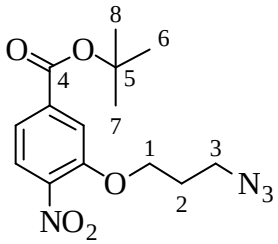
δ_H (500 MHz, $CDCl_3$) 11.48 (1H, s, H33 or H41), 11.43 (1H, s, H33 or H41), 8.46 (1H, t, *J* 5.2, H34), 8.37 (1H, d, *J* 8.8, Ar), 7.17-7.13 (2H, m, Ar), 7.05 (1H, dd, *J* 8.2; 1.6, Ar), 6.94 (1H, d, *J* 1.6, Ar), 6.66 (1H, d, *J* 8.2, Ar), 4.31 (2H, s br, H47), 4.11 (2H, t, *J* 6.2, H20), 4.07 (2H, t, *J* 5.8, H1), 3.83 (2H, t, *J* 6.8, H3), 3.78 (2H, d, *J* 7.2, H15), 3.70 (2H, qt, *J* 6.4, H22), 3.62 (2H, d, *J* 7.4, H24), 2.19 (2H, quint, *J* 6.5, H21), 2.15-2.06 (3H, m, H2 and H25), 1.97 (1H, sept, *J* 6.8, H16), 1.49 (9H, s, H30-23, or H38-40, or H44-46), 1.48 (18H, s, H7-9 and H11-13), 1.46 (9H, s, H30-32, or H38-40, or H44-46), 1.21 (9H, s, H30-32, or H38-40, or H44-46), 0.95 (6H, d, *J* 6.8, H26-27), 0.78 (6H, d, *J* 6.7, H17-H18). δ_C (125 MHz, $CDCl_3$) 175.5 (C23), 173.0 (C14), 163.5 (C35), 156.3 (C36 and C42), 154.3 (C19), 153.1 (C28), 152.6 (C4 and C5), 147.3 (Ar), 145.5 (Ar), 140.3 (Ar), 135.7 (Ar), 132.1 (Ar), 131.3 (Ar), 122.7 (Ar), 121.3 (Ar), 118.1 (Ar), 113.3 (Ar), 111.1 (Ar), 110.5 (Ar), 83.0 (C6 and C10 or C29 or C37 and C43), 82.6 (C6 and C10 or C29 or C37 and C43), 82.4 (C4 C6 and C10 or C29 or C37 and C43), 66.3 (C20), 66.2 (C1), 55.1 (C15), 53.0 (C24), 43.8 (C3), 37.6 (C22), 29.3 (C2), 28.8 (C21), 28.7 (C25), 28.5 (C16), 28.3 (C30-32 or C38-40 or C44-46), 28.3 (C30-32 or C38-40 or C44-46), 28.0 (C7-9 and C11-13), 27.5 (C30-32 or C38-40 or C44-46), 20.2 (C26-27), 19.9 (C17-18). IR υ_{max} 3366, 2963, 1724, 1639, 1619, 1368, 1330, 1133. HRMS (ESI) found 1099.6287 $C_{55}H_{87}N_8O_{15}^+$ $[M+H]^+$ requires 1099.6285.

3-(2-amino-5-(((2-(3-((amino(iminio)methyl)amino)propoxy)-4-(isobutylcarbamoyl)phenyl) carbamoyl)(isobutyl)carbamoyl)phenoxy)propan-1-aminium•2 TFA (1)



Following the general procedure for *N*-boc deprotection, protected benzoylurea **36** (79 mg, 0.081 mmol) was stirred for 1 hour in TFA to yield a brown oil (81 mg). A small amount was purified via semi-preparative HPLC (35 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (18 mg, 0.03 mmol) as a pale yellow oil: δ_H (500 MHz, MeOD) 8.32 (1H, d, *J* 8.4, Ar), 7.52 (1H, d, *J* 1.5, Ar), 7.50 (1H, dd, *J* 8.2; 1.8, Ar), 7.13 (1H, dd, *J* 8.0; 1.4, Ar), 7.10 (1H, d, *J* 1.5, Ar), 6.86 (1H, d, *J* 8.1, Ar), 4.24 (2H, t, *J* 5.4, H9), 4.19 (2H, t, *J* 5.5, H1), 3.85 (2H, d, *J* 7.2, H5), 3.58 (2H, t, *J* 7.1, H11), 3.26-3.19 (4H, m, H3 and H13), 2.26-2.15 (4H, m, H2 and H10), 1.92.010-1.91 (2H, m, H6 and H14), 0.99 (6H, d, *J* 6.7, H 15-16), 0.83 (6H, d, *J* 6.7, H7-8). δ_C (125 MHz, MeOD) 176.6 (C17), 168.6 (C19), 161.0 (q, *J* 36, F_3CCO_2), 157.7 (C20), 153.0 (C18), 148.0 (Ar), 145.8 (Ar), 141.2 (Ar), 131.0 (Ar), 129.9 (Ar), 124.0 (Ar), 123.3 (Ar), 120.2 (Ar), 119.0 (Ar), 114.1 (Ar), 111.4 (Ar), 110.3 (Ar), 65.8 (C1 and C9), 55.2 (C5), 47.5 (C13), 38.5 (C11), 37.5 (C3), 28.8 (C10), 28.7 (C6 and C14), 27.4 (C2), 19.6 (C15-16), 19.2 (C7-8). IR υ_{max} 3360, 3191, 2962, 1673, 1631, 1527, 1193, 1137. HRMS (ESI) found 599.3668 $C_{30}H_{47}N_8O_5^+$ $[M+H]^+$ requires 599.3664.

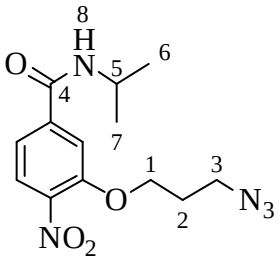
tert-Butyl 3-(3-azidopropoxy)-4-nitrobenzoate (44)



Carboxylic acid **7** (1.4 g, 5.26 mmol, 1.0 eq) was dissolved in DCM (30 mL) and cooled to 0 ºC and DMAP (96 mg, 0.79 mmol, 0.15 eq), *tert*-butanol (2 mL, 21 mmol, 4.0 eq), and *N,N'*-dicyclohexylcarbodiimide (1.25 g, 6 mmol, 1.15 eq) were added. The mixture was warmed to room temperature and stirred for 68 hours before 10 % NaOH

(100 mL) and DCM (100 mL) were added. The organic layer was washed with water (2 x 100 mL) and brine (1 x 100 mL), then dried ($MgSO_4$) and purified *via* flash column chromatography (10:1 petrol:ether) to yield the title compound as a yellow solid (952 mg, 2.95 mmol, 56 %): MP 74-75. δ_H (400 MHz, $CDCl_3$) 7.78 (1H, d, *J* 8.36, Ar), 7.67 (1H, d, *J* 1.3, Ar), 7.60 (1H, dd, *J* 8.3; 1.5, Ar), 4.23 (2H, t, *J* 5.8, H1), 3.65 (1H, t, *J* 6.5, H3), 2.07 (2H, quint, *J* 6.1, H2), 1.58 (9H, s, H 6-8). δ_C (100 MHz, $CDCl_3$) 163.6 (C4), 151.5 (Ar), 142.0 (Ar), 136.9 (Ar), 125.2 (Ar), 121.4 (Ar), 115.3 (Ar), 82.5 (C5), 66.1 (C1), 47.7 (C3), 28.4 (C2), 28.0 (C6-8). IR υ_{max} 3113, 3079, 2986, 2086, 1719, 1611, 1521, 1249. HRMS (ESI) found 345.1163 $C_{14}H_{18}N_4O_5Na^+$ $[M+Na]^+$ requires 345.1169.

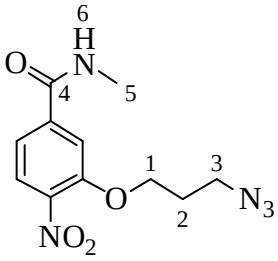
3-(3-Azidopropoxy)-*N*-isopropyl-4-nitrobenzamide (45)



Following the general procedure I for amidation, azido benzoic acid **7** (470 mg, 1.77 mmol, 1 eq) was coupled to isopropyl amine (125 mg, 2.12 mol, 1.3 eq) and purified (1:1 ether:petrol) to yield the title compound as a yellow oil (294 mg, 0.96 mmol, 54 %): δ_H (400 MHz, $CDCl_3$) 7.82 (1H, d, *J* 7.6, Ar), 7.57 (1H, d, *J* 1.8, Ar), 7.27 (1H, dd, *J* 8.3; 1.5, Ar), 6.18 (1H, d, *J* 7.1, H4), 4.30-4.19 (3H, m, H1 and H5), 3.56 (2H, t, *J* 5.1, H3), 2.08 (2H, quint, *J* 6.1, H2), 1.26 (6H, d, *J* 6.6, H6-7). δ_C (100 MHz, $CDCl_3$) 164.6 (C4), 152.1 (Ar), 141.1 (Ar), 140.3 (Ar), 125.7 (Ar), 117.9 (Ar), 114.0 (Ar), 66.2 (C1), 47.7 (C3), 42.5 (C5), 28.4 (C2), 22.7(C7-8). IR υ_{max} 3295, 3078, 2973, 2097, 1638, 1587, 1286. HRMS (ESI) found 330.1162 $C_{13}H_{17}N_5O_4Na^+$ $[M+Na]^+$ requires 330.1173.

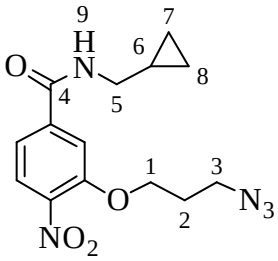
3-(3-Azidopropoxy)-*N*-methyl-4-nitrobenzamide (46)

Following the general procedure I for amidation, azido benzoic acid **7** (1.08 g, 4.83 mmol, 1.0 eq) was coupled to methyl amine (2M in THF, 6.13 mol, 1.5 eq) and purified (2 %



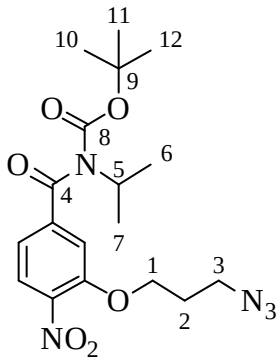
MeOH in DCM) to yield the title compound as a viscous orange oil (1.09 g, 4.64 mmol, 96 %): δ_H (400 MHz, $CDCl_3$) 8.08 (1H, d, *J* 8.1, Ar), 7.59 (1H, d, *J* 1.3, Ar), 7.28 (1H, dd, *J* 8.3; 1.3, Ar), 6.41 (1H, br s, H6), 4.25 (2H, t, *J* 5.8, H1), 3.57 (2H, t, *J* 6.3, H3), 3.03 (3H, d, *J* 4.5, H5), 2.09 (2H, quint, *J* 6.1, H2). δ_C (100 MHz, $CDCl_3$) 163.2 (C4), 152.1 (Ar), 141.2 (Ar), 139.7 (Ar), 125.8 (Ar), 117.9 (Ar), 114.0 (Ar), 66.2 (C1), 47.7 (C3), 48.0 (C6), 28.4 (C5), 27.1 (C2). IR υ_{max} 3315, 3087, 2945, 2100, 1650, 1524, 1255. HRMS (ESI) found 302.0855 $C_{11}H_{13}N_5O_4Na^+$ $[M+Na]^+$ requires 302.0860.

3-(3-Azidopropoxy)-*N*-(cyclopropylmethyl)-4-nitrobenzamide (47)



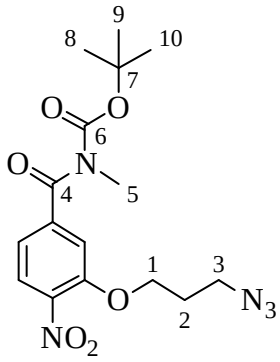
Following the general procedure I for amidation, azido benzoic acid **7** (800 mg, 3.0 mmol) was coupled to cyclopropylmethyl amine (0.31 mL, 3.60 mmol) and purified (1:1 ether:petrol) to yield the title compound as an orange oil (775 mg, 2.43 mmol, 81 %): δ_H (400 MHz, $CDCl_3$) 7.74 (1H, d, *J* 8.3, Ar), 7.53 (1H, s, Ar), 7.33 (1H, d, *J* 8.4, Ar), 7.09 (1H, t, *J* 5.3, H9), 4.14 (2H, t, *J* 5.8, H1), 3.48 (2H, t, *J* 6.4, H3), 3.22 (2H, dd, *J* 6.7, 5.9, H5), 1.99 (2H, quint, *J* 6.1, H2), 1.04-0.94 (1H, m, H6), 0.47 (2H, dd, *J* 13.2; 5.2, H7-8). 0.19 (2H, dd, *J* 10.3; 4.9, H7-8). δ_C (100 MHz, $CDCl_3$) 165.5 (C4), 151.9 (Ar), 140.9 (Ar), 140.1 (Ar), 125.5 (Ar), 118.4 (Ar), 113.8 (Ar), 66.2 (C1), 47.7 (C3), 45.2 (C5), 28.3 (C2), 10.6 (C6), 3.5 (C7-8). IR υ_{max} 3307, 2970, 2098, 1739, 1523, 1352, 1230 cm^{-1}. HRMS (ESI) found 342.1162 $C_{14}H_{17}N_5O_4Na^+$ $[M+Na]^+$ requires 342.1173.

***tert*-Butyl (3-(3-azidopropoxy)-4-nitrobenzoyl)(isopropyl)carbamate (48)**



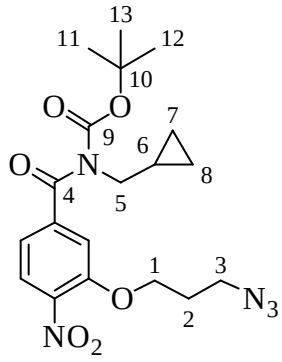
Following the general procedure for the Boc protection of amides, the azide **45** (424 mg, 1.41 mmol, 1 eq) was stirred for 5 days and purified (1:1 ether:petrol) to yield the title compound as a yellow oil (150 mg, 0.37 mmol, 26 %): δ_H (100 MHz, $CDCl_3$) 7.81 (1H, d, *J* 8.3, Ar), 7.23 (1H, d, *J* 1.4, Ar), 7.09 (1H, dd, *J* 8.3; 1.5, Ar), 4.68 (1H, sept, *J* 6.9, H5), 4.18 (2H, t, *J* 5.8, H1), 3.54 (2H, t, *J* 6.4, H3), 2.05 (2H, quint, *J* 6.1, H2), 1.39 (6H, d, *J* 6.9, H6 and H7), 1.19 (9H, s, H10-12). δ_C (100 MHz, $CDCl_3$) 171.0 (C4), 152.8 (C8), 152.0 (Ar), 143.8 (Ar), 140.7 (Ar), 125.4 (Ar), 118.8 (Ar), 113.3 (Ar), 84.7 (C9), 66.2 (C1), 49.0 (C5), 47.6 (C3), 28.4 (C2), 28.1 (C10-C12), 20.3 (C6 and C7). IR υ_{max} 2975, 2098, 1732, 1675, 1238, 1120. HRMS (ESI) found 430.1684 $C_{18}H_{25}N_5O_6Na^+$ $[M+Na]^+$ requires 430.1697.

***tert*-Butyl (3-(3-azidopropoxy)-4-nitrobenzoyl)(methyl)carbamate (49)**



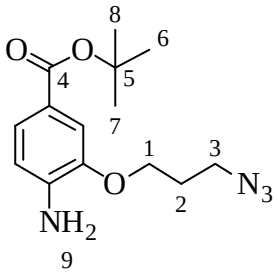
Following the general procedure for the Boc protection of amides, the azide **46** (1.09 g, 3.90 mmol) was stirred for 3 days and purified (5:2 petrol:ether) to yield the title compound as a yellow oil (1.07 g, 2.85 mmol, 73 %): δ_H (400 MHz, $CDCl_3$) 7.83 (1H, d, *J* 8.3, Ar), 7.17 (1H, d, *J* 1.3, Ar), 7.07 (1H, dd, *J* 8.3; 1.5, Ar), 4.18 (2H, t, *J* 5.8, H1), 3.55 (2H, t, *J* 6.4, H3), 3.28 (3H, s, H5), 2.06 (1H, sept, *J* 6.1, H2), 1.23 (9H, s, H8-10). δ_C (100 MHz, $CDCl_3$) 171.1 (C4), 152.8 (C6), 151.9 (Ar), 143.3 (Ar), 140.4 (Ar), 125.3 (Ar), 118.9 (Ar), 113.2 (Ar), 84.1 (C7), 66.2 (C1), 47.7 (C3), 32.4 (C5), 28.4 (C2), 27.5 (C8-C10). IR υ_{max} 2979, 2100, 1735, 1673, 1301, 1145. HRMS (ESI) found 402.1368 $C_{16}H_{21}N_5O_6Na^+$ $[M+Na]^+$ requires 402.1384.

***tert*-Butyl (3-(3-azidopropoxy)-4-nitrobenzoyl)(cyclopropylmethyl)carbamate (50)**



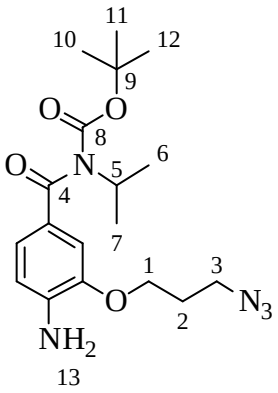
Following the general procedure for the Boc protection of amides, the azide **47** (881 mg, 2.43 mmol) was stirred for 5 days and purified (5:1 petrol:ether) to yield the title compound as a viscous yellow oil (881 mg, 2.10 mmol, 86 %.): δ_H (400 MHz, $CDCl_3$) 7.79 (1H, d, *J* 8.3, Ar), 7.18 (1H, s, Ar), 7.05 (1H, d, *J* 8.3, Ar), 4.16 (2H, t, *J* 5.8, H1), 3.62 (2H, d, *J* 7.2, H5), 3.50 (2H, t, *J* 6.4, H3), 2.02 (2H, quint, *J* 6.0, H2), 1.21-1.09 (1H, m, H6), 1.19 (9H, s, H11-13), 0.45 (2H, dd, *J* 12.6; 5.7, H7-8), 0.30 (2H, dd, *J* 10.2; 4.9, H7-8). δ_C (100 MHz, $CDCl_3$) 171.0 (C4), 153.0 (C9), 151.9 (Ar), 143.6 (Ar), 140.4 (Ar), 125.3 (Ar), 118.7 (Ar), 113.2 (Ar), 83.7 (C10), 66.2 (C1), 50.0 (C5), 47.6 (C3), 28.3 (C2). 27.4 (C11-13), 10.55 (C6), 3.52 (C7-8). IR υ_{max} 2979, 2881, 2098, 1732, 1672, 1524, 1344. HRMS (ESI) found 442.1679, $C_{19}H_{25}N_5O_6Na^+$ $[M+Na]^+$ requires 442.1697.

***tert*-Butyl 4-amino-3-(3-azidopropoxy)benzoate (51)**



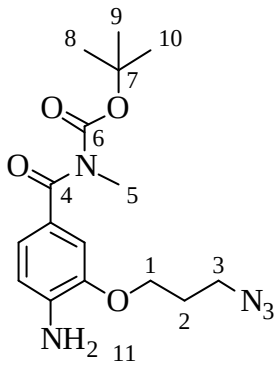
Following the general procedure for nitro reduction, the aromatic nitro compound **44** (952 mg, 2.95 mmol) was reduced and purified (3:2 petrol:ether) to yield the title compound as a colorless oil (560 mg, 1.92 mmol, 65 %): δ_H (250 MHz, $CDCl_3$) 7.48 (1H, dd, *J* 8.1; 1.7, Ar), 7.40 (1H, d, *J* 1.7, Ar), 6.62 (1H, d, *J* 8.2, Ar), 4.24 (2H, br s, H9), 4.08 (2H, t, *J* 6.0, H1), 3.46 (2H, t, *J* 6.6, H3), 2.04 (2H, quint, *J* 6.3, H2), 1.55 (9H, s, H6-8). δ_C (62.5 MHz, $CDCl_3$) 166.6 (C4), 145.4 (Ar), 141.4 (Ar), 124.5 (Ar), 121.7 (Ar), 113.7 (Ar), 112.6 (Ar), 80.6 (C5), 65.3 (C1), 48.8 (C3), 29.0 (C2), 28.7 (C6-8). IR υ_{max} 3488, 3369, 2976, 2932, 2098, 1691, 1617, 1297. HRMS (ESI) found 315.1425 $C_{14}H_{20}N_4O_3Na^+$ $[M+Na]^+$ requires 315.1428.

(4-amino-3-(3-azidopropoxy)benzoyl)(isopropyl)carbamate (52)



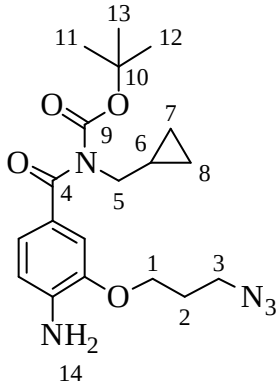
Following the general procedure for nitro reduction, azide **48** (232 mg, 0.57 mmol) was reduced and purified (3:2 pet:ether) to yield the title compound as a yellow oil (116 mg, 0.30 mmol, 52 %): δ_H (400 MHz, $CDCl_3$) 7.10 (1H, s, Ar), 7.06 (1H, dd, *J* 8.1, 1.7, Ar), 6.59 (1H, d, *J* 8.1, Ar), 4.51 (1H, sept, *J* 6.8, H5), 4.07 (2H, t, *J* 6.6, H1), 3.47 (2H, t, *J* 6.6, H3), 2.04 (2H, quint, *J* 6.3, H2), 2.00 (2H, br s, H13), 1.35 (6H, d, *J* 6.8, H6-7), 1.18 (9H, s, H10-12). δ_C (100 MHz, $CDCl_3$) 173.5 (C4), 153.8 (C8), 145.1 (Ar), 140.6 (Ar), 127.4 (Ar), 123.1 (Ar), 112.9 (Ar), 111.4 (Ar), 81.8 (C9), 65.1 (C1), 49.0 (C5), 48.3 (C3), 28.6 (C2), 27.6 (C10-12), 20.6 (C6-7). IR υ_{max} 3372, 2974, 2880, 2097, 1716, 1663, 1244. HRMS (ESI) found 400.1946, $C_{18}H_{27}N_5O_4Na^+$ $[M+Na]^+$ requires 400.1955.

***tert*-Butyl (4-amino-3-(3-azidopropoxy)benzoyl)(methyl)carbamate (53)**



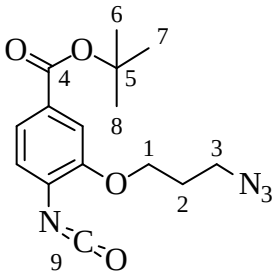
Following the general procedure for nitro reduction, azide **49** (462 mg, 1.22 mmol) was reduced and purified (3:2 ether:petrol) to yield the title compound as a tan oil (392 mg, 1.12 mmol, 92 %): δ_H (400 MHz, $CDCl_3$) 7.06-7.00 (2H, m, Ar), 6.59 (1H, d, *J* 8.0, Ar), 4.51 (1H, sept, *J* 6.8, H5), 4.16 (2H, br s, H11), 4.06 (2H, t, *J* 6.0, H1), 3.46 (2H, t, *J* 6.6, H3), 3.21 (3H, s, H5), 2.04 (2H, quint, *J* 6.3, H2), 1.22 (9H, s, H8-10). δ_C (100 MHz, $CDCl_3$) 173.5 (C4), 154.1 (C6), 145.0 (Ar), 140.4 (Ar), 126.2 (Ar), 123.2 (Ar), 112.9 (Ar), 111.6 (Ar), 82.3 (C7), 65.2 (C1), 48.3 (C3), 33.0 (C5), 28.6 (C2), 27.6 (C8-10). IR υ_{max} 3483, 3370, 2977, 2097, 1717, 1614, 1138. HRMS (ESI) found 372.1627, $C_{16}H_{23}N_5O_4Na^+$ $[M+Na]^+$ requires 372.1642.

***tert*-Butyl (4-amino-3-(3-azidopropoxy)benzoyl)(cyclopropylmethyl)carbamate (54)**



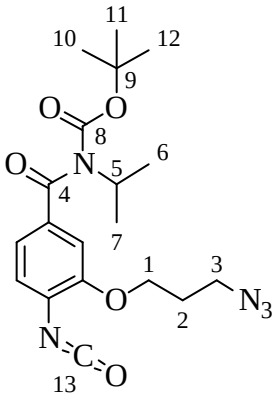
Following the general procedure for nitro reduction, azide **50** (881 mg, 2.10 mmol) was reduced and purified (1:1 petrol:ether) to yield the title compound as an orange oil (742 mg, 1.91 mmol, 91 %): δ_H (400 MHz, $CDCl_3$) 7.10 (1H, s, Ar), 7.06 (1H, dd, *J* 8.0, 1.7, Ar), 6.62 (1H, dd, *J* 8.0, 3.7, Ar), 4.21 (2H, br s, H14), 4.09 (2H, t, *J* 6.0, H1), 3.61 (2H, d, *J* 7.1, H5), 3.48 (2H, t, *J* 6.6, H3), 2.06 (2H, quint, *J* 6.3, H2), 1.27-1.21 (10H, m, H6 and H11-13), 0.45 (2H, dd, *J* 13.9; 4.5, H7-8), 0.30 (2H, dd, *J* 10.4; 4.8, H7-8). δ_C (100 MHz, $CDCl_3$) 173.5 (C4), 154.4 (C9), 145.1 (Ar), 140.3 (Ar), 126.8 (Ar), 122.9 (Ar), 111.4 (Ar), 82.1 (C10), 65.1 (C1), 50.5 (C5), 48.3 (C3), 28.6 (C2), 27.6 (C11-13), 10.9 (C6), 3.6 (C7-8). IR υ_{max} 3372, 2977, 2935, 2097, 1718, 1660, 1615, 1136. HRMS (ESI) found 412.1942, $C_{19}H_{27}N_5O_4Na^+$ $[M+Na]^+$ requires 412.1955.

***tert*-Butyl 3-(3-azidopropoxy)-4-isocyanatobenzoate (55)**



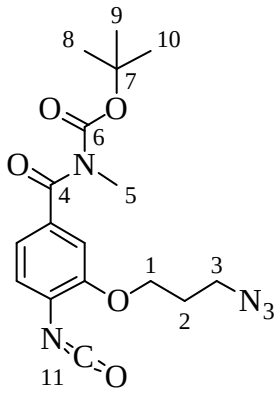
Following the general procedure for the formation of isocyanates, reaction with azide **51** (946 mg, 3.23 mmol) afforded the title compound as a tan oil (1.07 g, quant): δ_H (300 MHz, $CDCl_3$) 7.56-7.46 (2H, m, Ar), 6.97 (1H, d, *J* 7.4, Ar), 4.19 (2H, t, *J* 6.0, H1), 3.56 (2H, t, *J* 6.6, H3), 2.11 (2H, quint, *J* 6.3, H2), 1.57 (9H, s, H6-8). δ_C (75 MHz, $CDCl_3$) 164.9 (C4), 152.2 (Ar), 131.4 (Ar), 130.0 (Ar), 127.9 (C9), 123.1 (Ar), 122.8 (Ar), 111.8 (Ar), 81.4 (C5), 65.6 (C1), 47.8 (C3), 28.5 (C2), 28.1 (C6-8). IR υ_{max} 2978, 2935, 2240, 2097, 1708, 1292, 1159.

***tert*-Butyl (3-(3-azidopropoxy)-4-isocyanatobenzoyl)(isopropyl)carbamate (56)**



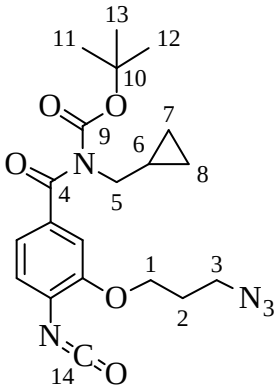
Following the general procedure for the formation of isocyanates, reaction with azide **52** (100 mg, 0.27 mmol) afforded the title compound as an off-white oil (108 mg, 0.27 mmol, quant): δ_H (300 MHz, $CDCl_3$) 7.09-6.90 (3H, m, Ar), 4.58 (1H, sept, *J* 9.2, H5), 4.11 (2H, t, *J* 7.7, H1), 3.51 (2H, t, *J* 8.7, H3), 2.06 (2H, quint, *J* 8.1, H2), 1.34 (6H, d, *J* 9.1, H6-7), 1.13 (9H, s, H10-12). δ_C (75 MHz, $CDCl_3$) 171.5 (C4), 152.4 (C8), 151.4 (Ar), 135.4 (Ar), 130.0 (C13), 125.8 (Ar), 122.0 (Ar), 119.7 (Ar), 109.6 (Ar), 81.9 (C9), 64.6 (C1), 48.1 (C5), 46.8 (C3), 27.5 (C2), 26.6 (C10-12), 19.5 (C6-7). IR υ_{max} 2975, 2935, 2883, 2243, 2098, 1730, 1671.

***tert*-Butyl (3-(3-azidopropoxy)-4-isocyanatobenzoyl)(methyl)carbamate (57)**



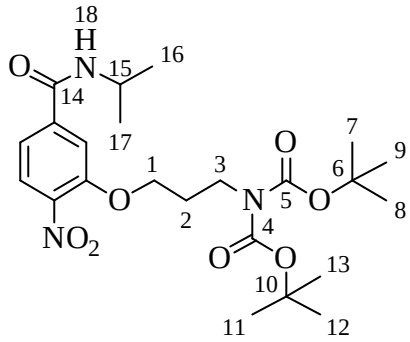
Following the general procedure for the formation of isocyanates, reaction with azide **53** (350 mg, 1.00 mmol) afforded the title compound as an off-white oil (375 mg, 1.00 mmol, quant.): δ_H (300 MHz, $CDCl_3$) 6.99-6.90 (3H, m, Ar), 4.10 (2H, t, *J* 7.7, H1), 3.50 (2H, t, *J* 8.7, H3), 3.20 (3H, s, H5), 2.05 (2H, quint, *J* 8.2, H2), 1.17 (9H, s, H8-10). δ_C (75 MHz, $CDCl_3$) 171.5 (C4), 152.5 (C6), 151.2 (Ar), 134.5 (Ar), 130.0 (C11), 125.5 (Ar), 121.9 (Ar), 119.8 (Ar), 109.6 (Ar), 82.2 (C7), 64.7 (C1), 46.8 (C3), 31.7 (C5), 27.5 (C2), 26.5 (C8-10). IR υ_{max} 2979, 2886, 2241, 2097, 1731, 1668, 1295, 1144.

***tert*-Butyl (3-(3-azidopropoxy)-4-isocyanatobenzoyl)(cyclopropylmethyl)carbamate (58)**



Following the general procedure for the formation of isocyanates, reaction with azide **54** (318 mg, 0.817 mmol) afforded the title compound as a yellow oil (328 mg, 0.78 mmol, 97 %): δ_H (500 MHz, $CDCl_3$) 7.10 (1H, d, *J* 1.6, Ar), 7.04 (1H, dd, *J* 7.9, 1.9, Ar), 6.97 (1H, d, *J* 8.2, Ar), 4.16 (2H, t, *J* 7.0, H1), 3.64 (2H, d, *J* 7.1, H5), 3.54 (2H, t, *J* 6.3, H3), 2.10 (2H, quint, *J* 6.3, H2), 1.27-1.21 (10H, m, H6 and H11-13), 0.47 (2H, dd, *J* 13.9; 4.5, H7-8), 0.33 (2H, dd, *J* 10.4; 4.8, H7-8). δ_C (125 MHz, $CDCl_3$) 172.5 (C4), 153.8 (C9), 135.9 (Ar), 131.0 (Ar), 126.5 (C14), 123.0 (Ar), 120.6 (Ar), 110.6 (Ar), 83.0 (C10), 65.6 (C1), 50.4 (C5), 47.8 (C3), 28.5 (C2), 27.5 (C11-13), 10.8 (C6), 3.6 (C7-8). IR υ_{max} 2978, 2242, 2098, 1730, 1670, 1358, 1146.

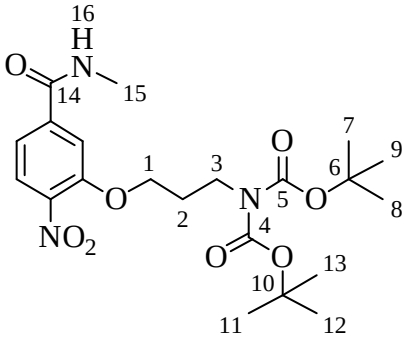
3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy- *N*-isopropyl-4-nitrobenzamide (59)



Following the general procedure II for amidation, di-Boc acid **31** (1.05 g, 2.38 mmol) was coupled with isopropyl amine (183 mg, 3.09 mmol, 1.3 eq) and purified (1:1 petrol:ether) to yield the title compound as a yellow oil (428 mg, 0.89 mmol, 37 %): δ_H (400 MHz, $CDCl_3$) 7.81 (1H, d, *J* 7.8, Ar), 7.54 (1H, d, *J* 1.3, Ar), 7.27 (1H, dd, *J* 8.4; 1.3, Ar), 6.08 (1H, d, *J* 7.6, H18), 4.31-4.22 (1H, m, H15), 4.19 (2H, t, *J* 6.1, H1), 3.79 (2H, t, *J* 6.9, H3), 2.12 (2H, quint, *J* 6.3, H2), 1.47 (18H, s, H7-9 and H11-13), 1.27 (6H, d, *J* 6.6, H16 and H17). δ_C (100 MHz, $CDCl_3$) 164.6 (C14), 152.5 (C4 and C5), 152.3 (Ar), 141.3 (Ar), 140.1 (Ar), 125.6 (Ar), 117.7 (Ar), 114.0 (Ar), 82.5 (C6 and C10), 67.8 (C1), 43.6 (C3), 42.4 (C15),

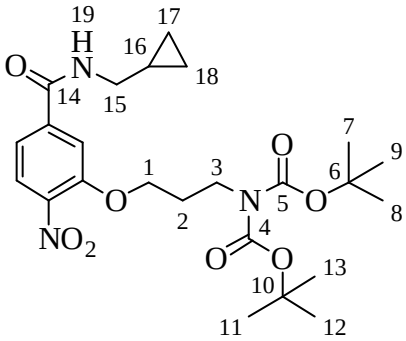
28.6 (C2), 28.0 (C7-9, C11-13), 22.3 (C16 and C17). IR υ_{max} 3336, 2977, 1775, 1642, 1528, 1366, 1111. HRMS (ESI) found 504.2313 $C_{23}H_{35}N_3O_8Na^+$ $[M+Na]^+$ requires 504.2316.

3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy- *N*-methyl-4-nitrobenzamide (60)



Following the general procedure II for amidation, di-Boc acid **31** (1.20 g, 2.72 mmol) was coupled with methyl amine (2 M in THF, 3.53 mmol, 1.3 eq) and purified (1:1 petrol:ether) to yield the title compound as a pale yellow solid (748 mg, 1.65 mmol, 61 %): MP 135-138. δ_H (400 MHz, MeOD) 7.87 (1H, d, *J* 8.4, Ar), 7.67 (1H, d, *J* 1.8, Ar), 7.48 (1H, dd, *J* 8.4; 1.6, Ar), 4.27 (2H, t, *J* 5.7, H1), 3.83 (2H, t, *J* 7.0, H3), 2.95 (3H, s, H15), 2.12 (2H, quint, *J* 6.3, H2), 1.48 (18H, s, H7-9 and H11-13). δ_C (100 MHz, MeOD) 167.4 (C14), 152.9 (C4 and C5), 152.0 (Ar), 142.0 (Ar), 139.6 (Ar), 125.3 (Ar), 119.0 (Ar), 113.7 (Ar), 82.8 (C6 and C10), 67.7 (C1), 44.0 (C3), 28.7 (C2), 28.0 (C7-9, C11-13), 26.1 (C15). IR υ_{max} 3385, 2977, 1766, 1657, 1521, 1368, 1112. HRMS (ESI) found 476.2002 $C_{21}H_{31}N_3O_8Na^+$ $[M+Na]^+$ requires 476.2003.

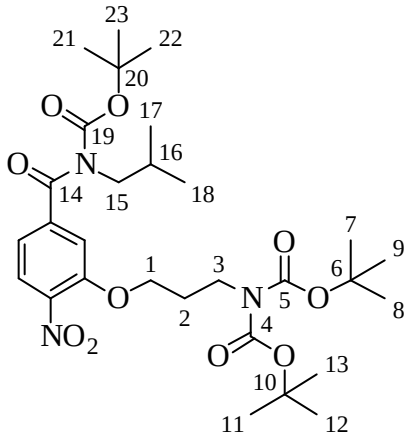
3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy- *N*-cyclopropylmethyl-4-nitrobenzamide (61)



Following the general procedure II for amidation, di-Boc acid **31** (300 mg, 0.68 mmol) was coupled with cyclopropylmethylamine (63 mg, 0.89 mmol) and purified (3:2 ether:petrol) to yield the title compound as a yellow oil (233 mg, 0.47 mmol, 69 %): δ_H (400 MHz, $CDCl_3$) 7.84 (1H, d, *J* 8.3, Ar), 7.58 (1H, s, Ar), 7.33 (1H, d, *J* 8.4, Ar), 6.29 (1H, s, H19), 4.21 (2H, t, *J* 6.0, H1), 3.80 (2H, t, *J* 7.0,

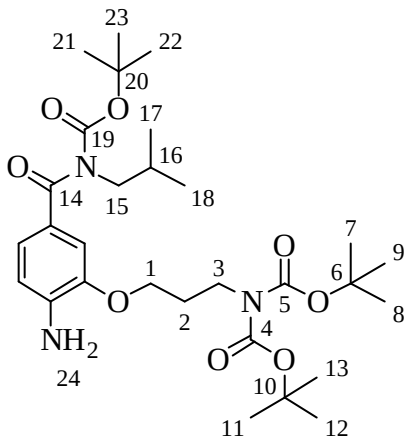
H3), 3.32 (2H, t, *J* 6.3, H15), 2.13 (2H, quint, *J* 6.4, H2), 1.50 (18H s, H7-9 and H11-13), 1.07 (1H, quint, *J* 3.8, H16), 0.57 (2H, dd, *J* 13.6; 4.8, H17-18), 0.29 (2H, dd, *J* 10.5; 5.0, H7-8). δ_C (100 MHz, $CDCl_3$) 165.3 (C14), 152.5 (C4 and C5), 152.3 (Ar), 141.3 (Ar), 139.9 (Ar), 125.6 (Ar), 117.9 (Ar), 114.0 (Ar), 82.5 (C6 and C10), 67.8 (C1), 45.3 (C15), 43.6 (C3), 28.6 (C2), 28.0 (C7-9 and C11-13), 10.6 (C16), 3.6 (C17-18). IR υ_{max} 3337, 2980, 1772, 1692, 1527, 1366. HRMS (ESI) found 516.2305 $C_{36}H_{49}N_7O_{12}Na^+$ $[M+Na]^+$ requires 516.2316.

3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy-*tert*-butyl-*N*-isobutyl-4-nitrobenzamide (62)



Following the general procedure for the Boc protection of amides, di-Boc amide **33** (760 mg, 1.53 mmol) was stirred for 3 days and purified (4:1 petrol:ether) to yield the title compound as a pale yellow oil (610 mg, 1.02 mmol, 67 %): δ_H (400 MHz, $CDCl_3$) 7.76 (1H, d, *J* 8.3, Ar), 7.13 (1H, d, *J* 1.8, Ar), 7.00 (1H, dd, *J* 8.3; 1.5, Ar), 4.09 (2H, t, *J* 6.1, H1), 3.72 (2H, t, *J* 7.3, H3), 3.57 (2H, d, *J* 7.3, H15), 2.10-1.95 (3H, m, H2 and H16), 1.41 (18H, s, H7-9 and H11-13), 1.18 (9H, s, H21-23), 0.88 (6H, d, *J* 8.0, H17-18). δ_C (100 MHz, $CDCl_3$) 170.9 (C14), 153.2 (Ar), 152.4 (C4-5 or C19), 152.1 (C4-5 or C19), 143.4 (Ar), 140.5 (Ar), 125.2 (Ar), 118.4 (Ar), 113.2 (Ar), 83.7 (C20), 82.5 (C6 or C10), 82.5 (C6 or C10), 67.8 (C1), 52.5 (C15), 43.5 (C3), 28.6 (C2), 28.1 (C16), 28.0 (C7-9 and C11-13), 27.4 (C21-23), 20.1 (C17 and C18). IR υ_{max} 2978, 2935, 1758, 1735, 1589, 1394, 1146. HRMS (ESI) found 618.2999 $C_{29}H_{45}N_3O_{10}Na^+$ $[M+Na]^+$ requires 618.2997.

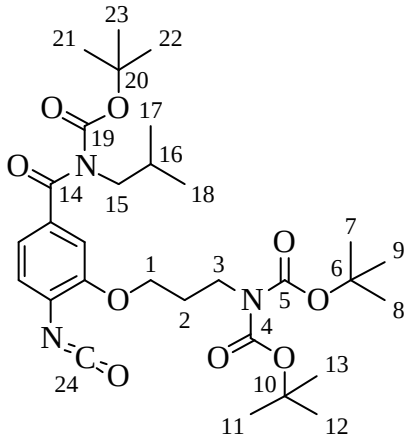
3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy-*tert*-butyl-*N*-isobutyl-4-aminobenzamide (63)



Palladium on activated carbon (55 mg, 5 % by mass palladium) was added to a stirred solution of the Boc-protected amide **62** (547 mg, 0.918 mmol) in EtOH (15 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 16 hours. The solution was filtered through Celite® with DCM (50 mL), concentrated, and purified *via* flash column chromatography (2:1 petrol:ether) to yield the title compound as a colorless oil (466 mg, 0.82 mmol, 90 %): δ_H (400 MHz, $CDCl_3$) 7.05 (1H, d, *J* 1.8, Ar), 7.05 (1H, dd, *J* 8.3; 1.5, Ar), 6.58 (1H, d, *J* 8.1, Ar), 4.22 (2H, br s, H24), 4.02 (2H, t, *J* 5.9, H1), 3.77 (2H, t, *J* 7.1, H3), 3.54 (2H, d, *J* 7.3, H15), 2.10-1.96 (3H, m, H2 and H16), 1.44 (18H, s, H7-9 and H11-13), 1.18 (9H, s, H21-23), 0.88 (6H, d, *J* 6.8, H17-18). δ_C (100 MHz, $CDCl_3$) 173.2 (C14), 154.6 (C4-5 or C19), 152.6 (C4-5 or C19), 145.4 (Ar),140.5 (Ar), 126.4 (Ar), 122.8 (Ar), 112.7 (Ar), 111.2 (Ar), 82.4 (C6 and C10), 81.9 (C20), 66.0 (C1), 53.0 (C15), 43.8 (C3), 28.8 (C2), 28.4 (C16), 28.0 (C7-9, C11-13), 27.6 (C21-23), 20.2 (C17 and C18). IR υ_{max} 3486, 3374, 2976, 1719, 1617, 1334, 1121. HRMS (ESI) found 588.3257 $C_{29}H_{47}N_3O_8Na^+$ $[M+Na]^+$ requires 588.3255.

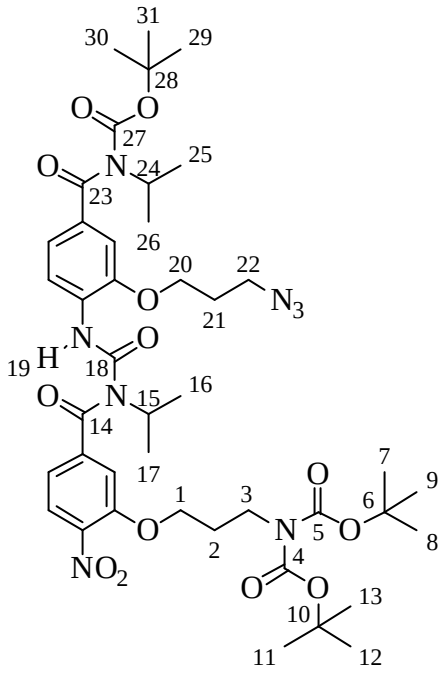
3-(3-(Bis(*tert*-butoxycarbonyl)amino)propoxy-*tert*-butyl-*N*-isobutyl-4-isocyanatobenzamide (64)

Following the general procedure for the formation of isocyanates, reaction with amine **63** (443 mg, 0.783 mmol) afforded the title compound as a tan oil (461 mg, quant): δ_H (500 MHz, $CDCl_3$) 7.05 (1H, d, *J* 1.3, Ar), 6.99 (1H, dd, *J* 8.1; 1.4, Ar), 6.92 (1H, d, *J* 8.1, Ar),



4.07 (2H, t, *J* 6.2, H1), 3.78 (2H, t, *J* 6.7, H3), 3.57 (2H, d, *J* 7.3, H15), 2.11 (2H, quint, *J* 6.5, H2), 2.04 (1H, sept, *J* 6.8, H16), 1.43 (18H, s, H7-9 and H11-13), 1.17 (9H, s, H21-23), 0.90 (6H, d, *J* 6.8, H17-18). δ_C (125 MHz, $CDCl_3$) 172.2 (C14), 154.0 (Ar), 152.6 (C4-5 or C19), 152.5 (C4-5 or C19),135.7 (Ar), 131.0 (Ar), 126.7 (C24), 122.8 (Ar), 120.4 (Ar), 110.6 (Ar), 82.9 (C6 and C10), 82.4 (C20), 67.0 (C1), 52.8 (C15), 43.2 (C3), 28.8 (C2), 28.3 (C16), 28.0 (C7-9, C11-13), 27.5 (C21-23), 20.2 (C17 and C18). IR υ_{max} 2977, 2243, 1786, 1713, 1422, 1121.

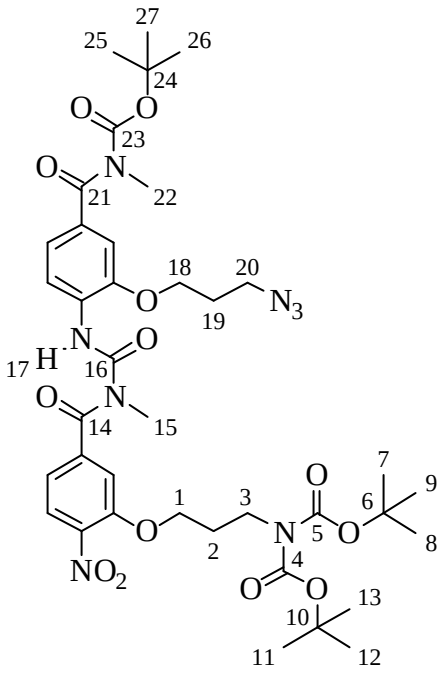
_N_,_N_-Di-_tert_-butyl 3-(3-azidopropoxy)-4-(3-(3-(3-((_tert_-butoxycarbonyl)amino)propoxy)-4-nitrobenzoyl)-3-isopropylureido)benzoate (65)



Following the general procedure for benzoylurea formation, amide **59** (134 mg, 0.28 mmol, 1.0 eq) was reacted with isocyanate **56** (110 mg, 0.31 mmol, 1.1 eq) and purified (1:1 petrol:ether) to yield the title compound as a yellow oil (117 mg, 0.13 mmol, 48 %): δ_H (400 MHz, $CDCl_3$) 11.27 (1H, s, H19), 8.32 (1H, d, *J* 8.8, Ar), 7.82 (1H, d, *J* 8.2, Ar), 7.27-7.10 (4H, m, Ar), 4.63 (1H, sept, *J* 6.8, H24), 4.19-4.07 (5H, m, H1, H20, and H15), 3.80 (2H, t, *J* 6.9, H22), 3.61 (2H, t, *J* 6.6, H3), 2.16-2.06 (4H, m, H2 and H21), 1.53 (6H, d, *J* 6.7, H16-17), 1.47 (18H, s, H7-9 and H11-13), 1.40 (6H, d, *J* 7.0, H25-26), 1.20 (9H, s, H29-31). δ_C (100 MHz, $CDCl_3$) 172.9 (C14 or C23), 172.7 (C14 or C23), 153.6 (Ar), 152.5 (C4-5), 152.5 (C27), 151.0 (Ar), 147.2 (C18), 141.9 (Ar), 140.6 (Ar), 133.5 (Ar), 130.6 (Ar), 126.1 (Ar),

121.4 (Ar), 118.6 (Ar), 117.1 (Ar), 112.0 (Ar), 110.3 (Ar), 82.5 (C6, C10, and C28), 68.0 (C1), 65.5 (C20), 53.4 (C15), 49.1 (C24), 48.0 (C22), 43.5 (C3), 28.6 (C21), 28.6 (C2), 28.0 (C7-9 and C11-13), 27.6 (C29-31), 21.0 (C16-17), 20.5 (C25-26). IR υ_{max} 2976, 2934, 2098, 1784, 1716, 1666, 1605, 1589. HRMS (ESI) found 907.4166 $C_{42}H_{60}N_8O_{13}Na^+$ $[M+Na]^+$ requires 907.4172.

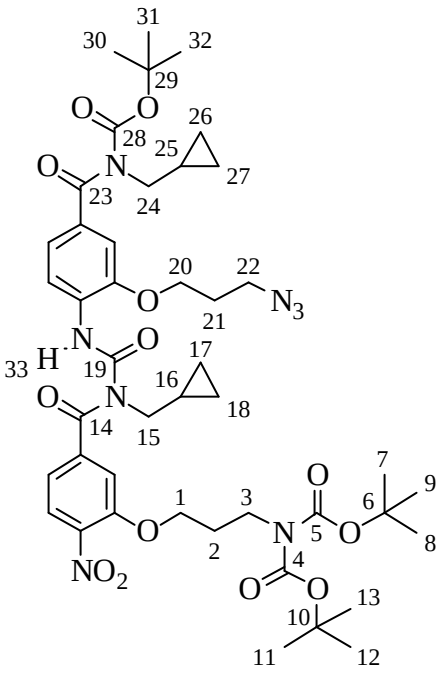
_N_,_N_-Di-*tert*-butyl (3-(3-azidopropoxy)-4-(3-(3-(3-((*tert*-butoxycarbonyl)amino) propoxy)-4-nitrobenzoyl)-3-methylureido)benzoyl)(methyl)carbamate (66)



Following the general procedure for benzoylurea formation, amide **60** (365 mg, 0.81 mmol, 1.0 eq) was reacted with isocyanate **57** (330 mg, 0.89 mmol, 1.1 eq) and purified (3:1 ether:petrol) to yield the title compound as an off-white solid (601 mg, 0.73 mmol, 90 %): MP 56-57. δ_H (400 MHz, $CDCl_3$) 11.79 (1H, s, H17), 8.37 (1H, d, *J* 8.2, Ar), 7.91 (1H, d, *J* 8.2, Ar), 7.27-7.16 (4H, m, Ar), 4.21-4.17 (4H, m, H1 and H18), 3.81 (2H, t, *J* 6.9, H3), 3.66 (2H, t, *J* 6.7, H20), 3.30 (3H, s, H15 or H22), 3.26 (3H, s, H15 or H22), 2.19-2.15 (4H, m, H2 and H19), 1.49 (18H, s, H7-9 and H11-13), 1.24 (9H, s, H25-27). δ_C (100 MHz, $CDCl_3$) 173.0 (C14 or C21), 172.5 (C14 or C21), 153.7 (Ar), 152.5 (C4-5), 152.5 (C23), 151.4 (Ar), 147.2 (C16), 140.9 (Ar), 140.7 (Ar), 132.7 (Ar), 130.3 (Ar), 126.0 (Ar), 121.5 (Ar), 118.6 (Ar), 118.0 (Ar), 112.9 (Ar), 110.4 (Ar), 83.0 (C24), 82.5 (C6 and C10), 68.0 (C1), 65.4 (C18), 48.0 (C20), 43.5 (C3), 35.5 (C15 or C22), 32.8 (C15 or C22), 28.7 (C19), 28.6 (C2), 28.0 (C7-9 and C11-13), 27.6 (C25-27). IR υ_{max} 3458, 2971, 2098, 1739,

1429, 1229, 1216. HRMS (ESI) found 851.3517 $C_{38}H_{52}N_8O_{13}Na^+$ $[M+Na]^+$ requires 851.3546.

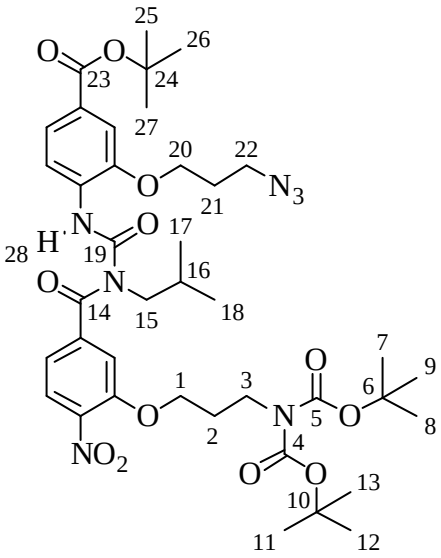
***tert*-Butyl(3-(3-azidopropoxy)-4-(3-(3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)-4-nitrobenzoyl)-3-(cyclopropylmethyl)ureido)benzoyl)(cyclopropylmethyl)carbamate (67)**



Following the general procedure for benzoylurea formation, amide **61** (354 mg, 0.72 mmol, 1.0 eq) was reacted with isocyanate **58** (328 mg, 0.79 mmol, 1.1 eq) and purified (1:1 ether:petrol) to yield the title compound as a white foam (507 mg, 0.56 mmol, 78 %): δ_H (400 MHz, $CDCl_3$) 11.73 (1H, s, H33), 8.36 (1H, d, *J* 8.8, Ar), 7.90 (1H, d, *J* 8.2, Ar), 7.20-7.14 (3H, m, Ar), 7.10 (1H, dd, *J* 8.2; 1.4, Ar), 4.17 (4H, t, *J* 5.7, H1 and H20), 3.80 (2H, t, *J* 6.9, H3), 3.70 (2H, d, *J* 6.8, H15), 3.68-3.62 (4H, m, H22 and H24), 2.19-2.09 (4H, m, H2 and H21), 1.47 (18H, s, H7-9 and H11-13), 1.27-1.23 (11H, m, H16, H25, H30-32), 0.53-0.43 (4H, m, H17-18 and H26-27), 0.35 (2H, dd, *J* 10.3; 4.8, H17-18 or H26-27), 0.14 (2H, dd, *J* 10.4; 4.7, H17-18 or H26-27). δ_C (100 MHz, $CDCl_3$) 172.9 (C14 or C23), 172.8 (C14 or C23), 154.0 (C19), 152.5 (C4 and C5 or C28), 152.4 (C4 and C5 or C28), 151.3 (Ar), 147.3 (Ar), 141.2 (Ar), 140.6 (Ar), 133.0 (Ar), 130.4 (Ar), 126.0 (Ar), 121.3 (Ar), 118.7 (Ar), 118.1 (Ar), 112.8 (Ar), 110.3 (Ar), 82.7 (C29), 82.5 (C6 and C10), 68.0 (C1), 65.4 (C20), 51.1 (C15), 50.4 (C24), 48.0 (C22), 43.4 (C3), 28.7 (C2), 28.6 (C21), 28.0 (C7-9 and C11-13), 27.6 (C30-32), 11.3 (C16), 10.8 (C25), 4.3 (C17-18), 3.6 (C26-27). IR υ_{max} 3219, 2979, 2099, 1716,

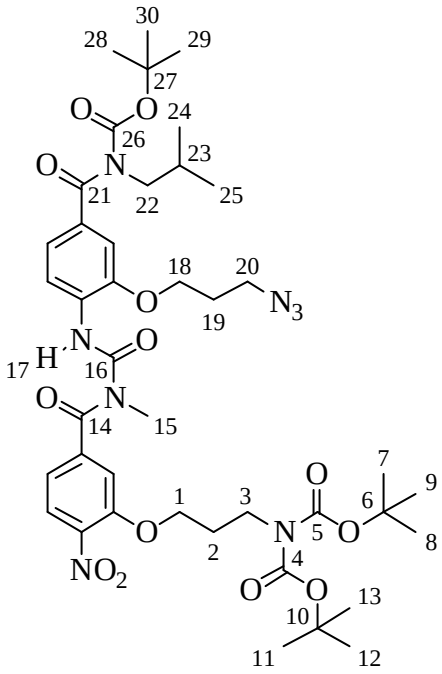
1532, 1314, 1122. HRMS (ESI) found 931.4139 $C_{44}H_{60}N_8O_{13}Na^+$ $[M+Na]^+$ requires 931.4172.

_N,N_-Di-_tert_-butyl-3-(3-azidopropoxy)-4-(3-(3-(3-((_tert_-butoxycarbonyl)amino)propoxy)-4-nitrobenzoyl)-3-isobutylureido)benzoate (68)



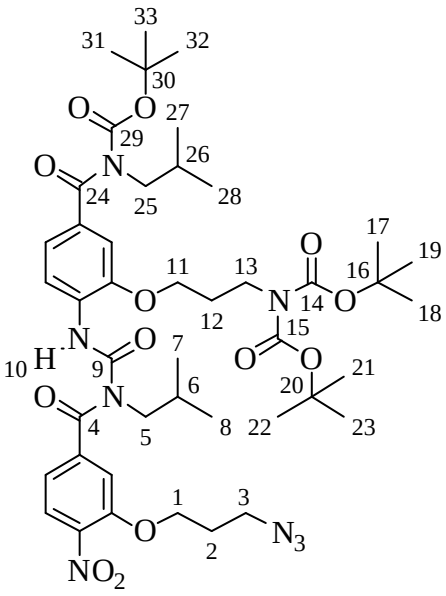
Following the general procedure for benzoylurea formation, amide **33** (255 mg, 0.51 mmol, 1.0 eq) was reacted with isocyanate **55** (175 mg, 0.55 mmol, 1.1 eq) and purified (3:1 petrol:ether) to yield the title compound as a white foam (380 mg, 0.47 mmol, 91 %): δ_H (500 MHz, $CDCl_3$) 11.66 (1H, s, H28), 8.37 (1H, d, *J* 8.5, Ar), 7.88 (1H, d, *J* 8.2, Ar), 7.63 (1H, dd, *J* 8.5; 1.5, Ar), 7.51 (1H, d, *J 1.9*, Ar), 7.26 (1H, s, Ar), 7.15 (1H, d, *J* 1.2, Ar), 7.09 (1H, dd, *J* 8.2; 1.4, Ar), 4.22-4.14 (4H, m, H1 and H20), 3.81 (2H, t, *J* 6.9, H3), 3.68-3.61 (4H, m, H15 and H22), 2.19-2.06 (4H, m, H2 and H21), 1.96 (1H, sept, *J* 6.8, H16), 1.58 (9H, s, H25-27), 1.47 (18H, s, H7-9 and H11-13), 0.80 (6H, d, *J* 6.7, H17-18). δ_C (100Hz, $CDCl_3$) 173.2 (C14), 165.4 (C23), 152.5 (C19), 152.4 (C4-5), 151.1 (Ar), 147.1 (Ar), 141.2 (Ar), 140.7 (Ar), 132.5 (Ar), 127.2 (Ar), 125.9 (Ar), 123.2 (Ar), 118.8 (Ar), 118.6 (Ar), 113.3 (Ar), 111.4 (Ar), 82.5 (C6 and C10), 81.0 (C24), 68.0 (C1 or C20), 65.3 (C1 or C20), 54.1 (C15), 48.0 (C22), 43.5 (C3), 28.8 (C2 or C21), 28.6 (C2 or C21), 28.6 (C16), 28.2 (C25-27), 28.0 (C7-9 and C11-13), 19.9 (C17-18). IR υ_{max} 3222, 2974, 2098, 1784, 1529, 1367, 1106. HRMS (ESI) found 836.3791 $C_{39}H_{55}N_7O_{12}Na^+$ $[M+Na]^+$ requires 836.3801.

***tert*-Butyl(3-(3-azidopropoxy)-4-(3-(3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)-4-nitrobenzoyl)-3-methylureido)benzoyl)(isobutyl)carbamate (69)**



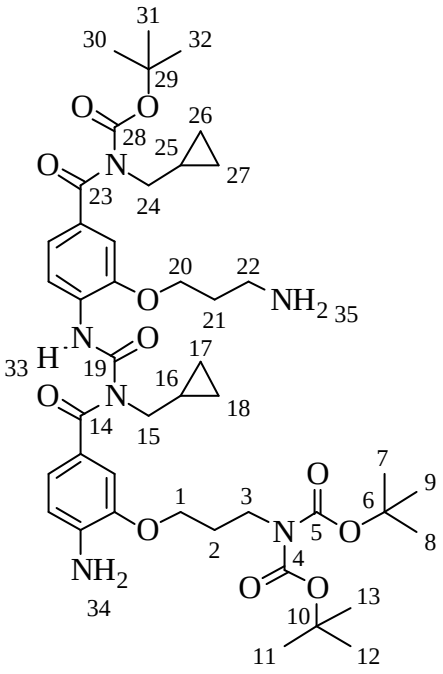
Following the general procedure for benzoylurea formation, amide **60** (154 mg, 0.34 mmol, 1.0 eq) was reacted with isocyanate **17** (149 mg, 0.36 mmol, 1.05 eq) and purified (1:1 petrol:ether) to yield the title compound as an off white oil (200 mg, 0.23 mmol, 64 %): δ_H (400 MHz, $CDCl_3$) 11.78 (1H, s, H17), 8.36 (1H, d, *J* 8.9, Ar), 7.91 (1H, d, *J* 8.2, Ar), 7.21-7.15 (3H, m, Ar), 7.12 (1H, dd, *J* 8.2; 1.4, Ar), 4.22-4.14 (4H, m, H1 and H18), 3.80 (2H, t, *J* 6.9, H3 or H20), 3.69-3.60 (4H, m, H22 and H3 or H20), 3.26 (3H, s, H15), 2.20-2.06 (5H, m, H2, H19, and H23), 1.48 (18H, s, H7-9 and H11-13), 1.23 (9H, s, H28-30), 0.96 (6H, d, *J* 6.7, H24-25). δ_C (100 MHz, $CDCl_3$) 172.8 (C14 or C21), 172.5 (C14 or C21), 154.2 (C4-5 or C26), 152.5 (C4-5 or C26), 152.5 (C16), 151.4 (Ar), 147.3 (Ar), 140.9 (Ar), 140.8 (Ar), 132.9 (Ar), 130.3 (Ar), 126.0 (Ar), 121.2 (Ar), 118.5 (Ar), 118.0 (Ar), 112.9 (Ar), 110.3 (Ar), 83.7 (C27), 82.5 (C6 and C10), 68.0 (C1 or C18), 65.3 (C1 or C18), 52.9 (C22), 48.0 (C3 or C20), 43.5 (C3 or C20), 35.5 (C15), 28.7 (C2 or C19), 28.5 (C2 or C19), 28.3 (C23), 28.0 (C7-9 and C11-13), 27.5 (C28-30), 20.2 (C24-25). IR υ_{max} 3220, 2976, 2099, 1784, 1716, 1534, 1367, 1336, 1120. HRMS (ESI) found 893.4003 $C_{41}H_{58}N_8O_{13}Na^+$ $[M+Na]^+$ requires 893.4016.

***tert*-Butyl (4-(3-(3-(3-azidopropoxy)-4-nitrobenzoyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl)carbamate (70)**



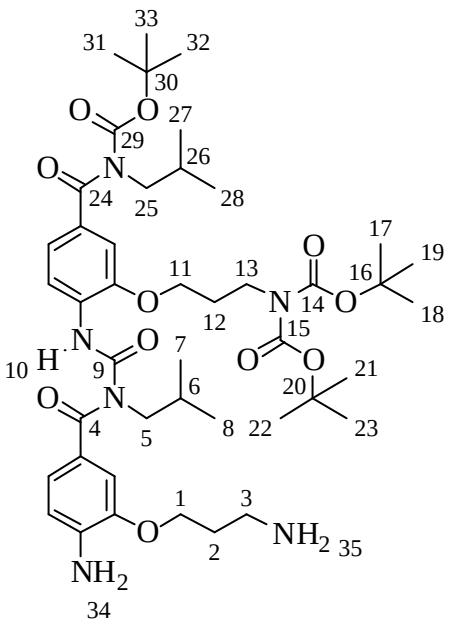
Following the general procedure for benzoylurea formation, amide **9** (265 mg, 0.83 mmol, 1.05 eq) was reacted with isocyanate **64** (463 mg, 0.78 mmol, 1.0 eq) and purified (2:1 petrol:ether) to yield the title compound as a yellow foam (609 mg, 0.67 mmol, 81 %): δ_H (400 MHz, $CDCl_3$) 11.33 (1H, s, H10), 8.31 (1H, d, *J* 8.3, Ar), 7.89 (1H, d, *J* 8.2, Ar), 7.19 (1H, d, *J* 1.2, Ar), 7.17-7.11 (2H, m, Ar), 4.21 (2H, t, *J* 5.8, H1 or H11), 4.09 (2H, t, *J* 6.3, H1 or H11), 3.83 (2H, t, *J* 6.7, H3 or H13), 3.67-3.55 (6H, m, H3 or H13 and H5 and H25), 2.22-2.04 (5H, m, H2, H12, and H6 or H26), 1.98 (1H, sept, *J* 6.8, H6 or H26), 1.43 (18H, s, H17-19 and H21-23), 1.22 (9H, s, H31-33), 0.94 (6H, d, *J* 6.8, H7-8 or H27-28), 0.83 (6H, d, *J* 6.7, H7-8 or H27-28). δ_C (100 MHz, $CDCl_3$) 172.7 (C4 or C24), 172.5 (C4 or C24), 154.2 (C14-15 or C29), 152.7 (C14-15 or C29), 152.1 (C9), 151.4 (Ar), 147.5 (Ar), 141.5 (Ar), 140.6 (Ar), 132.9 (Ar), 130.4 (Ar), 126.0 (Ar), 121.1 (Ar), 118.9 (Ar), 118.6 (Ar), 113.3 (Ar), 110.5 (Ar), 82.7 (C30), 82.4 (C16 and C20), 66.6 (C1 or C11), 66.4 (C1 or C11), 54.0 (C5 or C25), 53.0 (C5 or C25), 47.6 (C3 or C13), 43.2 (C3 or C13), 28.9 (C6 or C26), 28.6 (C6 or C26), 28.4 (C2 or C12), 28.3 (C2 or C12), 28.0 (C17-19 and C21-23), 27.6 (C31-33), 20.2 (C7-8 or C27-28), 19.9 (C7-8 or C27-28). IR υ_{max} 3223, 2964, 2100, 1715, 1530, 1367, 1335, 1121. HRMS (ESI) found 935.4509 $C_{44}H_{64}N_8O_{13}Na^+$ $[M+Na]^+$ requires 935.4485.

***tert*-Butyl(4-(3-(4-amino-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)-3-(cyclopropylmethyl)ureido)-3-(3-aminopropoxy)benzoyl)(cyclopropylmethyl) carbamate (71)**



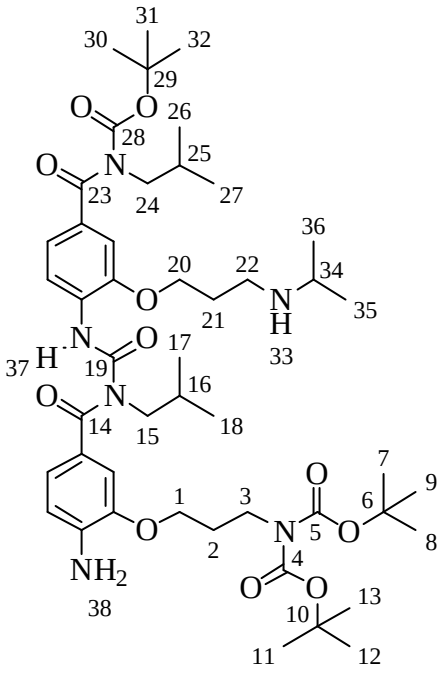
Palladium on activated carbon (200 mg, 10 % Pd by mass) was added to a stirred solution of benzoylurea **67** (102 mg, 0.112 mmol) in EtOAc (5 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen overnight. The solution was filtered through Celite® with EtOAc (50 mL) and the solvent removed *in vacuo* to yield the title compound as a pale tan oil (95 mg, quant): δ_H (300 MHz, $CDCl_3$) 11.47 (1H, s, H33), 8.36 (1H, d, *J* 8.7, Ar), 7.18-7.10 (2H, m, Ar), 7.01-6.89 (2H, m, Ar), 6.62 (1H, d, *J* 8.0, Ar), 4.30 (2H, s, H34), 4.13 (2H, t, *J* 5.9, H1 or H20), 4.04 (2H, t, *J* 5.5, H1 or H20), 3.86-3.77 (4H, m, H3 and H15 or H24), 3.64 (2H, d, *J* 7.1, H15 or H24), 2.94 (2H, t, *J* 6.3, H22), 2.12-1.96 (4H, m, H2 and H21), 1.86 (2H, br s, H35), 1.45 (18H, s, H7-9 and H11-13), 1.24-1.19 (11H, m, H16, H25, H30-32), 0.52-0.01 (8H, m, H17-18 and H26-27). δ_C (75 MHz, $CDCl_3$) 175.2 (C14 or C23), 173.3 (C14 or C23), 154.1 (C19), 152.8 (C4 and C5 or C28), 152.7 (C4 and C5 or C28), 147.6 (Ar), 145.6 (Ar), 140.2 (Ar), 132.3 (Ar), 131.4 (Ar), 124.3 (Ar), 122.2 (Ar), 121.2 (Ar), 118.4 (Ar), 113.3 (Ar), 110.9 (Ar), 110.6 (Ar), 82.6 (C29), 82.5 (C6 and C10), 67.0 (C1 or C20), 66.4 (C1 or C20), 52.3 (C15 or C24), 50.5 (C15 or C24), 43.8 (C3), 39.1 (C22), 33.0 (C2 or C21), 28.8 (C2 or C21), 28.1 (C7-9 and C11-13), 27.6 (C30-32), 11.4 (C16), 10.9 (C25), 4.0 (C17-18), 3.6 (C26-27). IR υ_{max} 3479, 3371, 3201, 2977, 1727, 1702, 1531, 1234, 1125. HRMS (ESI) found 853.4709 $C_{44}H_{65}N_6O_{11}^+$ $[M+H]^+$ requires 853.4706.

***tert*-Butyl(4-(3-(4-amino-3-(3-aminopropoxy)benzoyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl)carbamate (72)**



Palladium on activated carbon (20 mg, 5 % Pd by mass) was added to a stirred solution of benzoylurea **70** (96 mg, 0.105 mmol) in EtOAc (5 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen overnight. The solution was filtered through Celite® with EtOAc (50 mL) and the solvent removed *in vacuo* to yield the title compound as a tan foam (83 mg, 0.968 mmol, 92 %): δ_H (300 MHz, $CDCl_3$) 11.25 (1H, s, H10), 8.32 (1H, d, *J* 8.9, Ar), 7.13-7.07 (2H, m, Ar), 7.02-6.94 (2H, m, Ar), 6.63 (1H, d, *J* 7.0, Ar), 4.26 (2H, br s, H34), 4.09-4.01 (4H, m, H1 and H11), 3.82 (2H, t, *J* 6.6, H13), 3.75 (2H, d, *J* 7.1, H5 or H25), 3.58 (2H, d, *J* 7.3, H5 or H25), 2.89 (2H, t, *J* 6.8, H3), 2.17-1.89 (8H, m, H2, H6, H12, H26, and H35), 1.41 (18H, s, H17-19 and H21-23), 1.18 (9H, s, H31-33), 0.91 (6H, d, *J* 6.7, H7-8 or H27-28), 0.76 (6H, d, *J* 6.6, H7-8 or H27-28). δ_C (75 MHz, $CDCl_3$) 175.3 (C4 or C24), 172.9 (C4 or C24), 154.3 (C9), 152.7 (C14-15 or C29), 152.6 (C4 C14-15 or C29), 147.4 (Ar), 145.6 (Ar), 140.3 (Ar), 132.1 (Ar), 131.4 (Ar), 124.4 (Ar), 122.6 (Ar), 121.3 (Ar), 118.2 (Ar), 113.3 (Ar), 111.6 (Ar), 110.6 (Ar), 82.6 (C30), 82.3 (C16 and C20), 66.6 (C1 or C11), 66.5 (C1 or C11), 55.1 (C5 or C25), 53.0 (C5 or C25), 43.3 (C13), 39.2 (C3), 33.0 (C2), 29.0 (C12), 28.6 (C6 or C26), 28.4 (C6 or C26), 28.0 (C17-19 and C21-23), 27.6 (C31-33), 20.2 (C7-8 or C27-28), 20.0 (C7-8 or C27-28). IR υ_{max} 3369, 2962, 1730, 1698, 1496, 1317, 1188. HRMS (ESI) found 857.5023 $C_{44}H_{69}N_6O_{11}^+$ $[M+H]^+$ requires 857.5019.

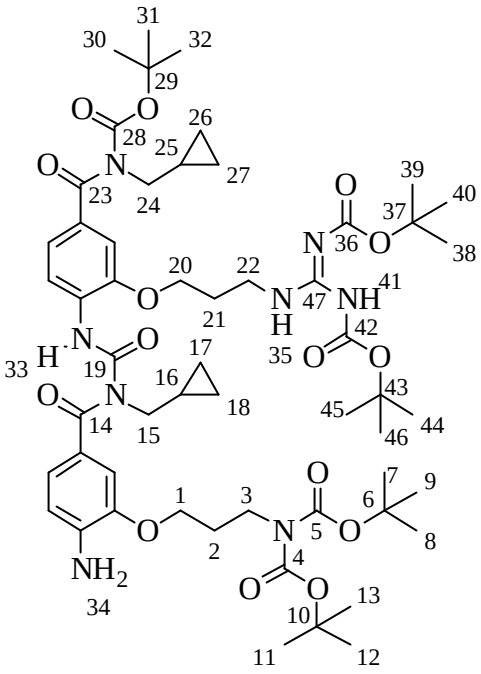
***tert*-Butyl(4-(3-(4-amino-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)-3-isobutylureido)-3-(3-(isopropylamino)propoxy)benzoyl)(isobutyl)carbamate (73)**



Palladium on activated carbon (36 mg, 5 % Pd by mass) was added to a stirred solution of benzoylurea **34** (36 mg, 0.039 mmol) in acetone (2 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen overnight. The solution was filtered through Celite® with EtOAc (30 mL) and the solvent removed *in vacuo* to yield the title compound as a tan oil (34 mg, 0.038 mmol, 96 %): δ_H (400 MHz, $CDCl_3$) 11.49 (1H, s, H37), 8.36 (1H, d, *J* 8.3, Ar), 7.15-7.11 (2H, m, Ar), 7.00 (1H, dd, *J* 8.1; 1.7, Ar), 6.92 (1H, d, *J* 1.6, Ar), 6.64 (1H, d, *J* 8.0, Ar), 4.31 (2H, br s, H38), 4.12 (2H, t, *J* 6.0, H1 or H20), 4.05 (2H, t, *J* 5.7, H1 or H20), 3.83 (2H, t, *J* 6.7, H3), 3.78 (2H, d, *J* 7.1, H15 or H24), 3.61 (2H, d, *J* 7.3, H15 or H24), 2.89 (2H, t, *J* 7.3, H22), 2.82 (1H, sept, *J* 6.4, H34), 2.15-1.88 (6H, m, H2, H16, H21, H25), 1.47 (18H, s, H7-9 and H11-13), 1.22 (9H, s, H30-32), 1.06 (6H, d, *J* 6.3, H35-36), 0.95 (6H, d, *J* 6.7, H17-18 or H26-27), 0.77 (6H, d, *J* 6.7, H17-18 or H26-27). δ_C (100 MHz, $CDCl_3$) 175.7 (C14 or C23), 173.1 (C14 or C23), 154.5 (C19), 152.8 (C4 and C5 or C28), 152.7 (C4 and C5 or C28), 147.6 (Ar), 145.6 (Ar), 140.4 (Ar), 132.2 (Ar), 131.4 (Ar), 124.4 (Ar), 122.7 (Ar), 121.3 (Ar), 118.3 (Ar), 113.4 (Ar), 111.2 (Ar), 110.6 (Ar), 82.7 (C29), 82.6 (C6 and C10), 67.2 (C1 or C20), 66.5 (C1 or C20), 55.3 (C15 or C24), 53.1 (C15 or C24), 48.9 (C34), 43.9 (C3), 43.9 (C22), 29.8 (C16 or C25), 28.9 (C16 or C25), 28.7 (C2 or C21), 28.5 (C2 or C21), 28.2 (C7-9 and C11-13), 27.7 (C30-32), 22.6 (C35-36), 20.3 (C17-18 or C26-27), 20.0 (C17-18 or C26-27). IR υ_{max}

3373, 2962, 1728, 1698, 1528, 1325, 1148. HRMS (ESI) found 889.5511 $C_{47}H_{75}N_6O_{11}^+$ $[M+H]^+$ requires 889.5488.

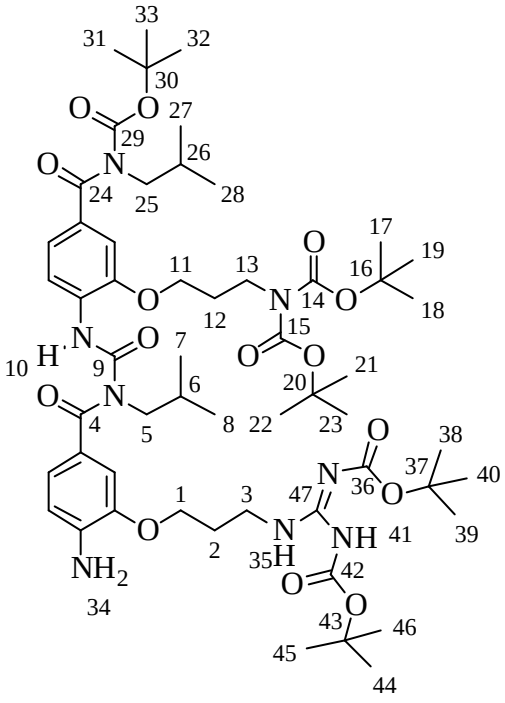
tert-Butyl(4-(3-(4-amino-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)-3-cyclopropylmethylureido)-3-(3-((2,2,10,10-tetramethyl-4,8-dioxo-3,9-dioxa-5,7-diazaundecan-6-yl)amino)propoxy)benzoyl)(cyclopropylmethyl)carbamate (74)



N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (42 mg, 0.134 mmol, 1.2 equiv) was added to a stirred solution of amine **71** (95 mg, 0.112 mmol, 1.0 eq) in DCM (3 mL). The reaction was stirred for 70 hours, before water (10 mL) and DCM (10 mL) were added. The mixture was washed with DCM (3 x 10 mL), dried (Na_2SO_4) and purified *via* flash column chromatography (1:1 petrol:ether) to yield the title compound as a tan oil (93 mg, 0.085 mmol, 76 %): δ_H (400 MHz, $CDCl_3$) 11.47 (1H, br s, H41), 11.44 (1H, s, H33), 8.44 (1H, t, *J* 5.4, H35), 8.36 (1H, d, *J* 8.3, Ar), 7.18-7.11 (2H, m, Ar), 7.00 (1H, dd, *J* 8.0; 1.6, Ar), 6.91 (1H, d, *J* 1.6, Ar), 6.63 (1H, d, *J* 8.0, Ar), 4.47 (2H, s br, H34), 4.08 (2H, t, *J* 6.1, H20), 4.04 (2H, t, *J* 5.7, H1), 3.86-3.62 (8H, m, H3, H15, H22, H24), 2.17 (2H, quint, *J* 6.3, H2 or H21), 2.08 (2H, quint, *J* 6.2, H2 or H21), 1.46 (9H, s, H38-40 or H44-46), 1.45 (18H, s, H7-9 and H11-13), 1.44 (10H, m, H38-40 or H44-46 and H16 or H25), 1.40 (1H, s, H16 or H25), 1.20 (9H, s, H30-32), 0.50-0.30 (8H, m, H17-18 and H26-27), 0.05 (2H, dd, *J* 10.3; 4.9, H17-18 or H26-27). δ_C (100 MHz, $CDCl_3$) 175.2 (C14 or C23), 173.3 (C14 or C23), 163.7 (C47), 156.4 (C19), 154.1 (C28 or C36 or C42), 153.2 (C28 or C36 or C42), 152.8 (C28 or C36 or C42), 152.7 (C4-5), 147.3 (Ar), 145.5 (Ar),

140.3 (Ar), 132.3 (Ar), 131.3 (Ar), 124.3 (Ar), 122.4 (Ar), 121.4 (Ar), 118.3 (Ar), 113.4 (Ar), 110.9 (Ar), 110.5 (Ar), 83.1 (C29 or C37 or C43), 82.7 (C29 or C37 or C43), 82.5 (C6 and C10), 79.3 (C29 or C37 or C43), 66.4 (C20), 66.3 (C1), 52.4 (C15 or C24), 50.5 (C15 or C24), 43.9 (C3), 37.7 (C22), 28.9 (C21), 28.8 (C2), 28.4 (C38-40 or C44-46), 28.1 (C38-40 or C44-46), 28.1 (C7-9 and C11-13), 27.6 (C30-32), 11.4 (C16 or C25), 10.6 (C16 or C25), 4.07 (C17-18 or C26-27), 3.64 (C17-18 or C26-27). IR υ_{max} 3340, 2979, 1721, 1683, 1493, 1364, 1334, 1128. HRMS (ESI) found 1117.5778 $C_{55}H_{82}N_8O_{15}Na^+$ $[M+Na]^+$ requires 1117.5792.

tert-Butyl (4-(3-(4-amino-3-(3-(bis(_tert_-butoxycarbonyl)guanidinopropoxy)benzoyl)-3-isobutylureido)-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl) carbamate (75)

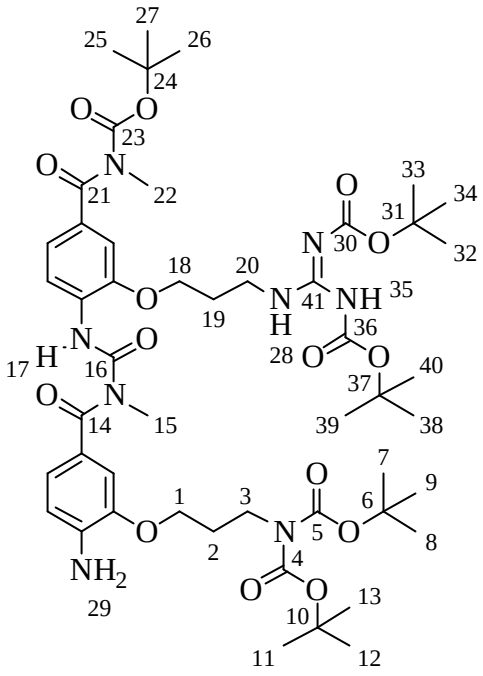


N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (36 mg, 0.116 mmol, 1.2 equiv) was added to a stirred solution of amine **72** (83 mg, 0.097 mmol, 1.0 eq) in DCM (3 mL). The reaction was stirred for 91 hours, before water (10 mL) and DCM (10 mL) were added. The mixture was washed with DCM (3 x 10 mL), dried (Na_2SO_4), filtered, and purified *via* flash column chromatography (3:2 ether:petrol) to yield the title compound as a colorless oil (104 mg, 0.095 mmol, 98 %): δ_H (400 MHz, $CDCl_3$) 11.52 (1H, s, H41), 11.32 (1H, s, H10), 8.48 (1H, t, *J* 4.8, H35), 8.34 (1H, d, *J* 8.9, Ar), 7.15-7.09 (2H, m, Ar), 7.03 (1H, dd, *J* 8.6; 1.9, Ar), 6.96 (1H, d, *J* 1.3, Ar), 6.65 (1H, d, *J* 8.1, Ar), 4.31 (2H, s br, H34), 4.13-4.03 (4H, m, H1 and H11), 3.84 (2H, t, *J* 6.7,

H13), 3.77 (2H, d, *J* 7.1, H5 or H25), 3.70-3.57 (4H, m, H3 and H15 or H25), 2.21-2.02 (5H, m, H2 and H12 and H6 or H26), 1.96 (1H, sept, *J* 6.8, H6 or H26), 1.48 (9H, s, H38-40 or H44-46), 1.47 (9H, s, H38-40 or H44-46), 1.43 (18H, s, H17-19 and H21-23), 1.20 (9H, s, H31-33), 0.94 (6H, d, *J* 6.7, H7-8 or H27-28), 0.78 (6H, d, *J* 6.7, H7-H8 or H27-28). δ_C (75 MHz, $CDCl_3$) 175.3 (C4 or C24), 173.0 (C4 or C24), 163.6 (C47), 156.3 (C9), 154.4 (C29 or C36 or C42), 153.5 (C29 or C36 or C42), 152.8 (C29 or C36 or C42), 152.7 (C14-15), 147.5 (Ar), 145.5 (Ar), 140.4 (Ar), 132.2 (Ar), 131.4 (Ar), 124.5 (Ar), 122.9 (Ar), 121.3 (Ar), 118.4 (Ar), 113.6 (Ar), 111.6 (Ar), 110.7 (Ar), 83.4 (C30 or C37 or C43), 82.7 (C30 or C37 or C43), 82.4 (C16 and C20), 79.5 (C30 or C37 or C43), 66.7 (C1 and C11), 55.2 (C5 or C25), 53.1 (C5 or C25), 43.4 (C13), 38.7 (C3), 29.1 (C2 or C12), 28.9 (C2 or C12), 28.7 (C6 or C26), 28.4 (C6 or C26), 28.4 (C38-40 or C44-46), 28.2 (C38-40 or C44-46), 28.1 (C17-19 and C21-23), 27.7 (C31-33), 20.3 (C7-8 or C27-28), 20.0 (C7-8 or C27-28). IR υ_{max} 3338, 2976, 1721, 1640, 1530, 1325, 1130. HRMS (ESI) found 1099.6288 $C_{55}H_{87}N_8O_{15}^+$ $[M+H]^+$ requires 1099.6285.

tert-Butyl (4-(3-(4-amino-3-(3-(bis(_tert_-butoxycarbonyl)guanidinopropoxy)benzoyl)-3-methylureido)-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)(methyl) carbamate (76)

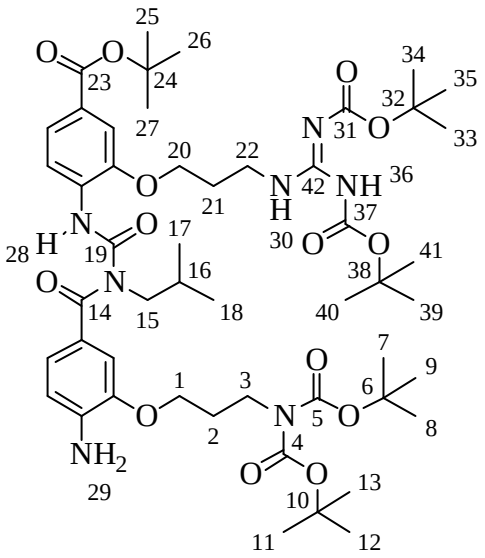
N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (10 mg, 0.031 mmol, 1.3 eq) and palladium on activated carbon (20 mg, 5 % Pd by mass) were added to a stirred solution of benzoylurea **66** (20 mg, 0.024 mmol, 1.0 eq) in EtOAc (1.5 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 42 hours. The solution was filtered through Celite® with EtOAc (30 mL), concentrated, and purified *via* flash column chromatography (3:2 ether:petrol) to yield the title compound as a colorless oil (17 mg, 0.017 mmol, 70 %): δ_H (500 MHz, $CDCl_3$) 11.83 (1H, s, H17), 11.48 (1H, br s,



H35), 8.51 (1H, t, *J* 4.9, H28), 8.37 (1H, d, *J* 8.4, Ar), 7.15 (1H, dd, *J* 8.4; 1.3, Ar), 7.07 (1H, s, Ar), 7.07 (1H, dd, *J* 8.1; 1.3, Ar), 6.97 (1H, s, Ar), 6.66 (1H, d, *J* 8.1, Ar), 4.13 (2H, t, *J* 6.0, H18), 4.08 (2H, t, *J* 5.7, H1), 3.83 (2H, t, *J* 6.8, H3), 3.73 (2H, q, *J* 6.2, H20), 3.34 (3H, s, H15 or H22), 3.28 (3H, s, H15 or H22), 2.21 (2H, quint, *J* 6.2, H19), 2.11 (2H, quint, *J* 6.2, H2), 1.48 (9H, s, H32-34 or H38-40), 1.47 (18H, s, H7-8 and H11-13), 1.45 (9H, s, H32-34 or H38-40), 1.20 (9H, s, H25-27). δ_C (125 MHz, $CDCl_3$) 174.8 (C14 or C21), 173.4 (C14 or C21), 163.5 (C41), 156.3 (C16), 153.9 (C23 or C30 or C36), 153.2 (C23 or C30 or C36), 153.0 (C23 or C30 or C36), 152.7 (C4-5), 147.3 (Ar), 145.4 (Ar), 140.4 (Ar), 132.0 (Ar), 131.3 (Ar), 123.8 (Ar), 122.5 (Ar), 121.6 (Ar), 118.3 (Ar), 113.2 (Ar), 111.0 (Ar), 110.6 (Ar), 83.0 (C24 or C31 or C37), 82.9 (C24 or C31 or C37), 82.5 (C6 and C10), 79.4 (C24 or C31 or C37), 66.3 (C1) 66.3 (C18), 43.9 (C3), 37.8 (C20), 36.7 (C15 or C22), 32.9 (C15 or C22), 28.8 (C2), 28.8 (C19), 28.3 (C25-27 or C32-34 or C38-40), 28.1 (C7-9 and C11-13), 27.6 (C25-27 or C32-34 or C38-40) 27.6 (C25-27 or C32-34 or C38-40). IR υ_{max} 3366, 3208, 2979, 1724, 1670, 1324, 1132. HRMS (ESI) found 1015.5362 $C_{49}H_{75}N_8O_{15}^+$ $[M+H]^+$ requires 1015.5346.

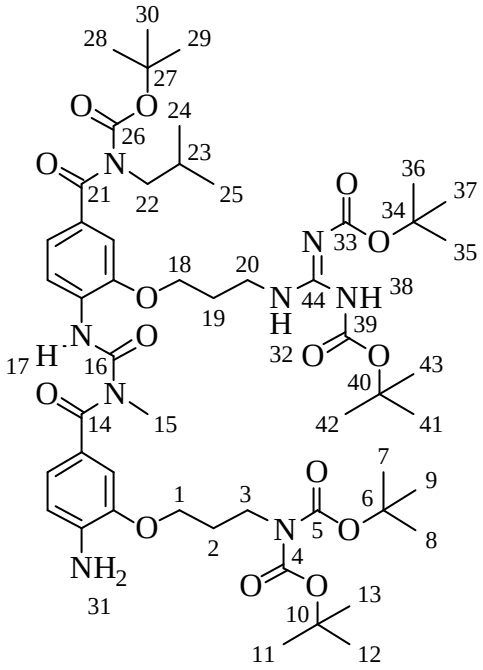
***tert*-Butyl 4-(3-(4-amino-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)-3-isobutylureido)-3-(3-(2,3-bis(*tert*-butoxycarbonyl)guanidino)propoxy)benzoate (77)**

N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (23 mg, 0.075 mmol, 1.2 eq) and palladium on activated carbon (51 mg, 5 % Pd by mass) were added to a stirred solution of benzoylurea **68** (51 mg, 0.063 mmol, 1.0 eq) in EtOAc (4.0 mL). The solution was flushed three times



with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 92 hours. The solution was filtered through Celite® with EtOAc (50 mL), concentrated, and purified *via* flash column chromatography (1:1 ether:petrol) to yield the title compound as a white foam (49 mg, 0.049 mmol, 78 %): δ_H (400 MHz, $CDCl_3$) 11.47 (1H, br s, H36), 11.37 (1H, s, H28), 8.47 (1H, br s, H30), 8.37 (1H, d, *J* 8.5, Ar), 7.60 (1H, dd, *J* 8.5; 1.6, Ar), 7.47 (1H, d, *J 1.6*, Ar), 7.04 (1H, dd, *J* 8.1; 1.6, Ar), 6.93 (1H, d, *J* 1.6, Ar), 6.65 (1H, d, *J* 8.1, Ar), 4.32 (2H, br s, H29), 4.13 (2H, t, *J* 6.1, H20), 4.06 (2H, t, *J* 5.7, H1), 3.83 (2H, t, *J* 6.8, H3), 3.78 (2H, d, *J* 7.2, H15), 3.70 (2H, q, *J* 5.9, H22), 2.19 (2H, quint, *J* 6.4, H21), 2.10 (2H, quint, *J* 6.3, H2), 1.97 (1H, sept, *J* 6.8, H16), 1.58 (9H, s, H25-27), 1.48 (9H, s, H33-35 or H39-41), 1.47 (18H, s, H7-9 and H11-13), 1.45 (9H, s, H33-35 or H39-41), 0.78 (6H, d, *J* 6.7, H17-18). δ_C (100Hz, $CDCl_3$) 175.5 (C14), 165.7 (C42), 156.3 (C23), 153.3 (C31 or C37), 152.8 (C4-5), 152.7 (C31 or C37), 147.2 (C19), 145.6 (Ar), 140.4 (Ar), 132.5 (Ar), 126.5 (Ar), 125.6 (Ar), 124.3 (Ar), 123.3 (Ar), 122.9 (Ar), 118.5 (Ar), 113.4 (Ar), 111.7 (Ar), 111.3 (Ar), 83.1 (C32 or C38), 82.5 (C6 and C10), 80.9 (C24), 79.3 (C32 or C38), 66.5 (C1), 66.3 (C20), 55.2 (C15), 43.8 (C3), 37.8 (C22), 28.9 (C2), 28.9 (C21), 28.7 (C16), 28.4 (C33-35 or C39-41), 28.4 (C33-35 or C39-41), 28.2 (C25-27 and C7-9 and C11-13), 20.1 (C17-18). IR υ_{max} 3337, 2977, 1703, 1638, 1619, 1368, 1164, 1133. HRMS (ESI) found 1000.5636 $C_{50}H_{78}N_7O_{14}^+$ $[M+H]^+$ requires 1000.5601.

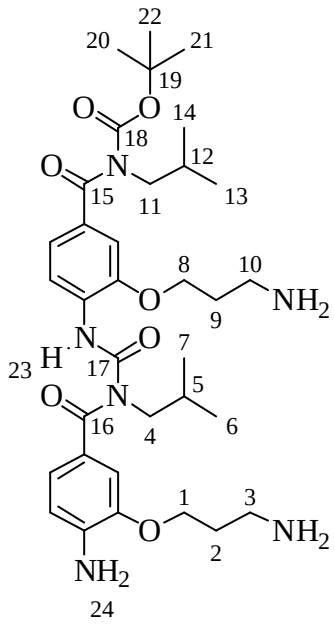
***tert*-Butyl (4-(3-(4-amino-3-(3-(bis(*tert*-butoxycarbonyl)guanidinopropoxy)benzoyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)(methyl) carbamate (78)**



N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (21 mg, 0.066 mmol, 1.2 eq) and palladium on activated carbon (48 mg, 5 % Pd by mass) were added to a stirred solution of benzoylurea **69** (48 mg, 0.055 mmol, 1.0 eq) in EtOAc (4.0 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 88 hours. The solution was filtered through Celite® with EtOAc (50 mL), concentrated, and purified *via* flash column chromatography (3:2 ether:petrol) to yield title compound as a white foam (52 mg, 0.049 mmol, 89 %): δ_H (400 MHz, $CDCl_3$) 11.81 (1H, s, H17), 11,48 (1H, br s, H38), 8.47 (1H, t, *J* 5.4, H32), 8.36 (1H, d, *J* 8.8, Ar), 7.16-7.13 (2H, m, Ar), 7.06 (1H, dd, *J* 8.1; 1.7, Ar), 6.97 (1H, d, *J* 1.6, Ar), 6.65 (1H, d, *J* 8.1, Ar), 4.61 (2H, br s, H31), 4.12 (2H, t, *J* 6.0, H18), 4.07 (2H, t, *J* 5.7, H1), 3.82 (2H, t, *J* 6.8, H3), 3.71 (2H, q, *J* 5.9, H20), 3.62 (2H, d, *J* 7.3 H22), 3.33 (3H, s, H15), 2.19 (2H, quint, *J* 6.3, H19), 2.15-2.06 (3H, m, H2 and H23), 1.47 (9H, s, H35-37 or H41-43), 1.47 (18H, s, H7-9 and H11-13), 1.41 (9H, s, H35-37 or H41-43), 1.23 (9H, s, H28-30), 0.95 (6H, d, *J* 6.7, H24-25). δ_C (100 MHz, $CDCl_3$) 174.8 (C14 or C21), 173.1 (C14 or C21), 163.7 (C44), 156.4 (C16), 153.2 (C26 or C33 and C39), 153.0 (C26 or C33 and C39), 152.8 (C4-5), 147.5 (Ar), 145.5 (Ar), 140.5 (Ar), 132.3 (Ar), 131.3 (Ar), 123.9 (Ar), 122.5 (Ar), 121.4 (Ar), 118.4 (Ar), 113.3 (Ar), 111.1 (Ar), 110.7 (Ar), 83.1 (C27 or C34 or C40), 82.7 (C27 or C34 or C40), 82.5 (C6 and C10), 79.3 (C27 or C34 or C40), 66.4 (C1 and C18),

53.0 (C22), 43.9 (C3), 37.8 (C20), 36.7 (C15), 28.9 (C19), 28.8 (C2), 28.8 (C23), 28.4 (C28-30 or C35-37 or C41-43), 28.4 (C28-30 or C35-37 or C41-43), 28.0 (C7-9 and C11-13), 27.6 (C28-30 or C35-37 or C41-43), 27.6 (C24-25). IR υ_{max} 3368, 2977, 1724, 1638, 1536, 1330, 1134. HRMS (ESI) found 1057.5823 $C_{52}H_{81}N_8O_{15}^+$ $[M+H]^+$ requires 1057.5816.

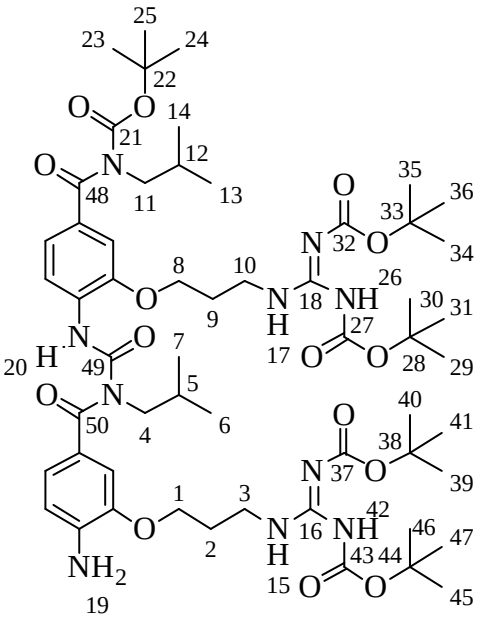
***tert*-Butyl (4-(3-(4-amino-3-(3-aminopropoxy)benzoyl)-3-isobutylureido)-3-(3-aminopropoxy) benzoyl)(isobutyl)carbamate (80)**



Following the general procedure for catalytic hydrogenation, benzoylurea **19** (495 mg, 0.67 mmol) was hydrogenated overnight to yield the title compound as a viscous green oil (439 mg, 0.67 mmol, *quant*), which was used for subsequent reactions without further purification: δ_H (500 MHz, $CDCl_3$) 11.39 (1H, s, H23), 8.36 (1H, d, *J* 8.4, Ar), 7.16 (1H, d, *J* 1.3, Ar), 7.13 (1H, dd, *J* 8.4; 1.5, Ar), 7.04-6.99 (2H, m, Ar), 6.66 (1H, d, *J* 7.9, Ar), 4.24 (2H, s, H24), 4.15-4.08 (4H, m, H1 and H8), 3.78 (2H, d, *J* 7.1, H4 or H11), 3.61 (2H, d, *J* 7.3, H4 or H11), 2.98-2.90 (4H, m, H3 or H10), 2.10 (1H, quint, *J* 6.9, H5 or H12), 2.04-1.94 (5H, m, H2 and H9 and H5 or H12), 1.21 (9H, s, H20-22), 0.95 (6H, d, *J* 6.7, H6 and H7 or H13 and H14), 0.79 (6H, d, *J* 6.7, H6 and H7 or H13 and H14). δ_C (125 MHz, $CDCl_3$) 175.9 (C15 or C16), 173.4 (C15 or C16), 154.8 (C18), 153.0 (C17), 147.9 (Ar), 146.1 (Ar), 140.6 (Ar), 132.6 (Ar), 131.7 (Ar), 124.9 (Ar), 122.9 (Ar), 121.6 (Ar), 118.6 (Ar), 113.8 (Ar), 112.0 (Ar), 111.0 (Ar), 83.1 (C19), 67.3 (C1 or C8), 66.9 (C1 or C8), 55.6 (C4 or C11), 53.4 (C4 or C11), 39.6 (C3 or C10), 39.5 (C3 or C10), 33.3 (C5 or C12), 33.2 (C5 or C12), 29.0 (C2 or C9), 28.8 (C2 or C9), 28.0 (C20-22), 20.7 (C6-7 or C13-14), 20.4 (C6-7 or C13-14). IR υ_{max} 3361, 3200,

2960, 1705, 1587, 1527, 1327, 1234. HRMS (ESI) found 657.3953 $C_{34}H_{53}N_6O_7^+$ $[M+H]^+$ requires 657.3970.

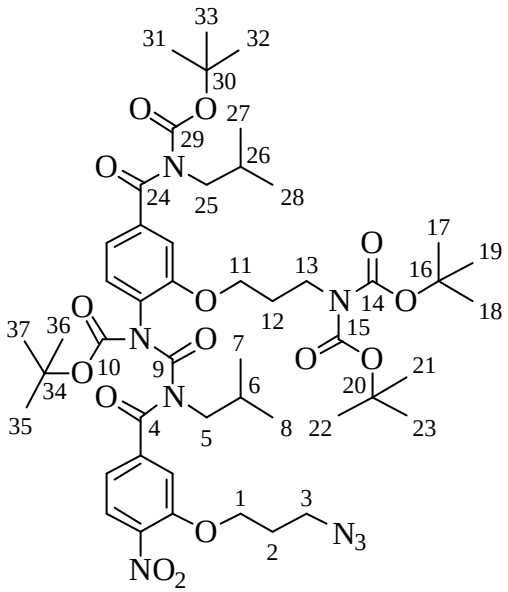
***tert*-Butyl (4-(3-(4-amino-3-(2,3-bis(*tert*-butoxycarbonyl)guanidino)propoxy)benzoyl)-3-isobutylureido)-3-(3-aminopropoxy) benzoyl)(isobutyl)carbamate (81)**



Following the general procedure for protected guanidino group formation, di-amine **80** (352 mg, 0.53 mmol, 1.0 eq) was reacted with *N*,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine (381 mg, 1.232 mmol, 2.3 eq) and purified (3:2 ether:petrol) to yield the title compound (351 mg, 0.31 mmol, 58 %) as a yellow oil: δ_H (400 MHz, $CDCl_3$) 11.52 (1H, s, H26 or H42), 11.47 (1H, s, H26 or H42), 11.40 (1H, s, H20), 8.50-8.42 (2H, m, H15 and H17), 8.35 (1H, d, *J* 8.8, Ar), 7.15-7.10 (2H, m, Ar), 7.05 (1H, dd, *J* 8.1; 1.3, Ar), 6.96 (1H, s, Ar), 6.65 (1H, d, *J* 8.1, Ar), 4.35 (2H, br s, H19), 4.13-4.06 (4H, m, H1 and H8), 3.77 (2H, d, *J* 7.1, H4), 3.72-3.62 (4H, m, H3 and H10), 3.60 (2H, d, *J* 7.3, H11), 2.22-2.05 (5H, m, H2 and H9 and H12), 1.96 (1H, sept, *J* 6.8, H5), 1.48 (9H, s, H29-31 or H34-36 or H39-41 or H45-47), 1.47 (18H, s, two of: H29-31, H34-36, H39-41, H45-47), 1.44 (9H, s, H29-31 or H34-36 or H39-41 or H45-47), 1.21 (9H, s, H23-25), 0.93 (6H, d, *J* 6.8, H13-14), 0.78 (6H, d, *J* 6.7, H6-7). δ_C (100 MHz, $CDCl_3$) 175.5 (C48), 173.0 (C50), 163.7 (C16 or C18), 163.6 (C16 or C18), 156.4 (C21 or C27 or C32 or C37 or C43), 156.2 (C21 or C27 or C32 or C37 or C43), 154.4 (C49), 153.4 (C21 or C27 or C32 or C37 or C43), 153.2 (C21 or C27 or C32 or C37 or C43), 152.6 (C21 or C27 or C32 or C37 or C43), 147.3 (Ar), 145.4 (Ar), 140.4 (Ar), 132.2 (Ar), 131.4 (Ar), 124.3 (Ar), 123.0 (Ar), 121.4 (Ar), 118.2 (Ar), 113.5 (Ar), 111.6 (Ar),

110.6 (Ar), 83.3 (C22 or C27 or C32 or C37 or C43), 83.1 (C22 or C28 or C33 or C38 or C44), 82.7 (C22 or C28 or C33 or C38 or C44), 79.4 (C22 or C28 or C33 or C38 or C44), 79.3 (C22 or C27 or C32 or C37 or C43), 66.7 (C1 or C8), 66.3 (C1 or C8), 55.2 (C4), 53.0 (C11), 38.6 (C3 or C10), 37.7 (C3 or C10), 28.9 (C2 or C9), 28.8 (C2 or C9), 28.6 (C5), 28.4 (C23-25 or C29-31 or C34-36 or C39-41 or C45-47), 28.2 (C23-25 or C29-31 or C34-36 or C39-41 or C45-47), 28.1 (C23-25 or C29-31 or C34-36 or C39-41 or C45-47), 28.1 (C12), 27.6 (C23-25 or C29-31 or C34-36 or C39-41 or C45-47), 20.3 (C13-14), 20.0 (C6-7). IR υ_{max} 3332, 2966, 2873, 1720, 1587, 1131. HRMS (ESI) found 1163.6372 $C_{56}H_{88}N_{10}O_{15}Na^{+}$ $[M+Na]^{+}$ requires 1163.6323.

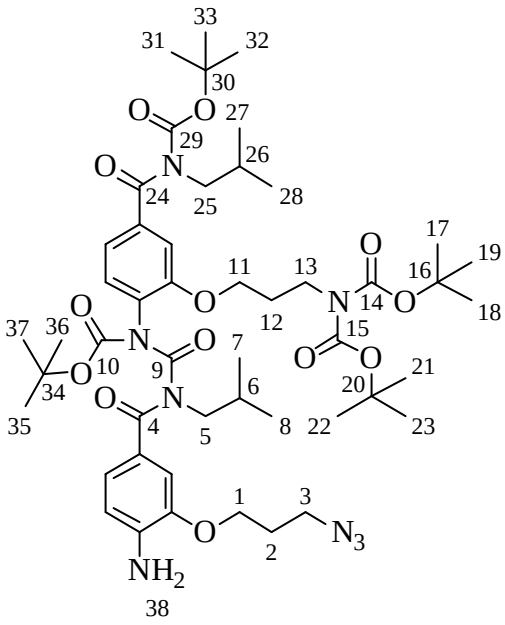
***tert*-Butyl (4-(3-(3-(3-azidopropoxy)-4-nitrobenzoyl)-1-(*tert*-butoxycarbonyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl) (isobutyl)carbamate (82)**



Di-*tert*-butyl dicarbonate (227 mg, 1.04 mmol, 2.0 eq) was added to a stirred solution of benzoylurea **70** (475 mg, 0.52 mmol, 1.0 eq) in THF (11 mL). DMAP (127 mg, 1.04 mmol, 2.0 eq) was added and the mixture stirred for 19 hours at room temperature. The reaction was diluted with NH_4Cl (50 mL) and DCM (50 mL) and extracted with DCM (3 x 50 mL). The organic layers were combined, dried (Na_2SO_4), filtered, concentrated, and purified *via* flash column chromatography (3:1 petrol:ether) to yield the title compound as a yellow foam (454 mg, 0.45 mmol, 86 %): δ_H (400 MHz, $CDCl_3$) 7.86 (1H, d, *J* 8.3, Ar), 7.38 (1H, d, *J* 1.5, Ar), 7.32 (1H, dd, *J* 8.3; 1.5, Ar), 7.08

(1H, d, *J* 1.8, Ar), 6.98 (1H, dd, *J* 8.1; 1.8, Ar), 6.58 (1H, d, *J* 7.6, Ar), 4.18 (2H, t, *J* 5.8, H1 or H11), 3.98 (2H, t, *J* 7.3, H1 or H11), 3.66 (2H, t, *J* 7.1, H3 or H13), 3.60-3.55 (6H, m, H3 or H13 and H5 and H25), 2.22 (1H, sept, *J* 6.8, H6 or H26), 2.13-1.94 (5H, m, H2, H12, and H6 or H26), 1.48 (18H, s, H17-19 and H21-23), 1.31 (9H, s, H35-37), 1.17 (9H, s, H31-33), 0.95 (6H, d, *J* 6.8, H7-8 or H27-28), 0.92 (6H, d, *J* 6.6, H7-8 or H27-28). δ_C (100 MHz, $CDCl_3$) 172.1 (C4 or C24), 169.0 (C4 or C24), 155.5 (C10), 154.3 (C14-15 or C29), 153.8 (C14-15 or C29), 152.6 (C9), 151.8 (Ar), 151.1 (Ar), 141.3 (Ar), 141.0 (Ar), 139.3 (Ar), 129.8 (Ar), 127.1 (Ar), 125.3 (Ar), 119.9 (Ar), 119.6 (Ar), 114.7 (Ar), 111.6 (Ar), 84.1 (C34), 83.2 (C30), 82.5 (C16 and C20), 67.0 (C1 or C11), 66.2 (C1 or C11), 53.4 (C5 or C25), 52.8 (C5 or C25), 47.6 (C3 or C13), 43.2 (C3 or C13), 28.8 (C6 or C26), 28.3 (C6 or C26), 28.3 (C2 or C12), 28.1 (C2 or C12), 28.0 (C17-19 and C21-23), 27.6 (C31-33), 27.4 (C35-37), 20.3 (C7-8 or C27-28), 20.1 (C7-8 or C27-28). IR υ_{max} 3128, 2978, 2098, 1731, 1672, 1335, 1152. HRMS (ESI) found 1035.5006 $C_{49}H_{72}N_8O_{15}Na^+$ $[M+Na]^+$ requires 1035.5009.

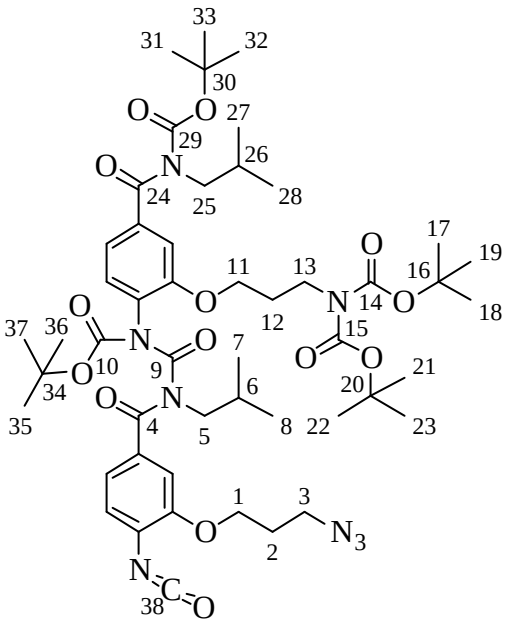
***tert*-Butyl(4-(3-(4-amino-3-(3-azidopropoxy)benzoyl)-1-(*tert*-butoxycarbonyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl) carbamate (83)**



Following the general procedure for nitro group reduction, benzoylurea **82** (454 mg, 0.45 mmol) was reduced and purified (2:1 petrol:ether) to yield the title compound as a white foam (182 mg, 0.18 mmol, 40 %): δ_H (400 MHz, $CDCl_3$) 7.18 (1H, dd, *J* 8.1; 1.5, Ar), 7.14 (1H, d, *J* 1.2, Ar), 7.06 (1H, d, *J* 1.5, Ar), 6.96 (1H, dd, *J* 7.5; 1.5, Ar), 6.71 (1H,

d, *J* 4.8, Ar), 6.59 (1H, d, *J* 7.5, Ar), 4.09 (2H, t, *J* 6.1, H1 or H11), 3.95 (2H, t, *J* 7.3, H1 or H11), 3.71-3.56 (6H, m, H3 or H13 and H5 and H25), 3.51 (2H, t, *J* 6.3, H3 or H13), 2.19 (1H, sept, *J* 6.8, H6 or H26), 2.14-1.89 (5H, m, H2, H12, and H6 or H26), 1.49 (18H, s, H17-19 and H21-23), 1.28 (9H, s, H35-37), 1.19 (9H, s, H31-33), 0.94 (12H, d, *J* 7.2, H7-8 and H27-28). δ_C (100 MHz, $CDCl_3$) 172.4 (C4 or C24), 171.6 (C4 or C24), 156.1 (C10), 154.4 (C14-15 or C29), 154.4 (C14-15 or C29), 154.0 (C9), 152.5 (Ar), 152.5 (Ar), 151.1 (Ar), 138.7 (Ar), 138.7 (Ar), 130.6 (Ar), 127.6 (Ar), 123.6 (Ar), 123.6 (Ar), 119.5 (Ar), 112.3 (Ar), 111.5 (Ar), 83.2 (C30 or C34), 83.2 (C30 or C34), 82.5 (C16 and C20), 67.0 (C1 or C11), 65.2 (C1 or C11), 55.5 (C5 or C25), 52.8 (C5 or C25), 48.3 (C3 or C13), 43.3 (C3 or C13), 28.9 (C6 or C26), 28.6 (C6 or C26), 28.3 (C2 or C12), 28.2 (C2 or C12), 28.1 (C17-19 and C21-23), 27.7 (C31-33), 27.5 (C35-37), 20.4 (C7-8 or C27-28), 20.2 (C7-8 or C27-28). IR υ_{max} 3373, 2964, 2099, 1733, 1663, 1367, 1150, 1124. HRMS (ESI) found 1005.5224 $C_{49}H_{74}N_8O_{13}Na^+$ $[M+Na]^+$ requires 1005.5268.

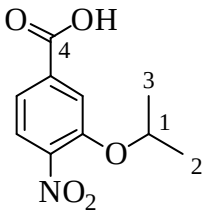
tert-Butyl(4-(3-(3-(3-azidopropoxy)-4-isocyanatobenzoyl)-1-(_tert_-butoxycarbonyl)-3-isobutylureido)-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl) carbamate (84)



Following the general procedure for the formation of isocyanates, reaction with amine **84** (182 mg, 0.19 mmol) afforded the title compound as a yellow foam (187 mg, quant): δ_H (300 MHz, $CDCl_3$) 7.24-7.18 (2H, m, Ar), 7.06 (1H, d, *J* 1.3, Ar), 7.02 (1H, d, *J* 7.9, Ar), 6.96 (1H, dd, *J* 8.0; 1.5, Ar), 6.48 (1H, d, *J* 8.0, Ar), 4.15 (2H, t, *J* 5.7, H1 or H11), 3.96 (2H, t, *J* 7.0, H1 or H11), 3.72-

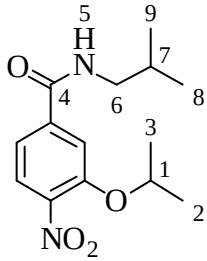
3.50 (8H, m, H3, H5, H13, and H25), 2.29-1.91 (6H, m, H2, H6, H12, and H26), 1.48 (18H, s, H17-19 and H21-23), 1.27 (9H, s, H35-37), 1.18 (9H, s, H31-33), 0.99-0.88 (12H, m, H7-8 and H27-28). δ_C (75 MHz, $CDCl_3$) 172.3 (C4 or C24), 171.0 (C4 or C24), 155.8 (C10), 154.4 (C14-15 or C29), 153.9 (C14-15 or C29), 152.6 (C9), 152.5 (Ar), 151.1 (Ar), 139.2 (Ar), 134.1 (Ar), 131.1 (Ar), 130.3 (Ar), 127.4 (C38), 127.2 (Ar), 122.9 (Ar), 121.7 (Ar), 119.6 (Ar), 111.8 (Ar), 111.6 (Ar), 83.7 (C30 or C34), 83.2 (C30 or C34), 82.5 (C16 and C20), 67.0 (C1 or C11), 65.7 (C1 or C11), 55.5 (C5 or C25), 52.9 (C5 or C25), 47.8 (C3 or C13), 43.3 (C3 or C13), 28.9 (C6 or C26), 28.5 (C6 or C26), 28.3 (C2 or C12), 28.2 (C2 or C12), 28.1 (C17-19 and C21-23), 27.8 (C31-33), 27.5 (C35-37), 20.4 (C7-8 or C27-28), 20.2 (C7-8 or C27-28). IR υ_{max} 2964, 2245, 2099, 1734, 1668, 1394, 1150, 1125.

3-Isopropoxy-4-nitrobenzoic acid (85)



Following the general procedure for S_NAr coupling, isopropyl alcohol (1 mL, 13.0 mmol, 1.2 eq) was added to acid **6** (2 g, 10.8 mmol, 1.0 eq) to yield the title compound as a yellow solid (2.56 g, quant): MP 155-157. δ_H (400 MHz, $CDCl_3$) 7.80-7.77 (2H, m, Ar), 7.73 (1H, dd, *J* 8.3; 1.3, Ar), 4.78 (1H, sept, *J* 6.0, H1), 1.42 (6H, d, *J* 6.1, H2-3). δ_C (100 MHz, $CDCl_3$) 170.0 (C4), 150.1 (Ar), 144.4 (Ar), 133.3 (Ar), 125.2 (Ar), 121.8 (Ar), 117.3 (Ar), 73.1 (C1), 21.8 (C2-C3). IR υ_{max} 2923, 1689, 1605, 1523, 1297. HRMS (ESI) found 248.0521 $C_{10}H_{11}NO_5$ Na^+ $[M+Na]^+$ requires 248.0529.

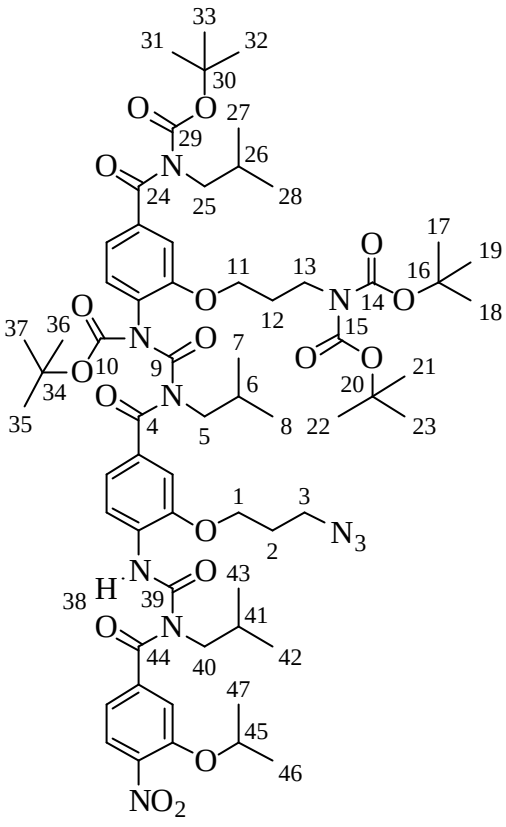
***N*-isobutyl-3-isopropoxy-4-nitrobenzamide (86)**



Following the general procedure I for amidation, isopropoxy benzoic acid **86** (2.56 g, 11.37 mmol, 1.0 eq) was coupled to isobutyl amine (1.25 g, 17.05 mmol, 1.5 eq) and purified (1:1 ether:petrol) to yield the title

compound as a viscous yellow oil (2.12 g, 7.57 mmol, 67 %): δ_H (400 MHz, $CDCl_3$) 7.73 (1H, d, *J* 8.3, Ar), 7.60 (1H, d, *J* 1.6, Ar), 7.20 (1H, dd, *J* 8.3; 1.7, Ar), 6.20 (1H, br s, H5), 4.80 (1H, sept, *J* 6.1, H1), 3.30 (2H, t, *J* 6.1, H6), 1.91 (1H, sept, *J* 6.7, H7), 1.40 (6H, d, *J* 6.0, H2-3), 0.99 (6H, d, *J* 6.7, H8-9). δ_C (100 MHz, $CDCl_3$) 165.4 (C4), 152.6 (Ar), 141.6 (Ar), 140.6 (Ar), 126.2 (Ar), 118.1 (Ar), 114.5 (Ar), 66.6 (C1), 48.0 (C6), 29.7 (C7), 28.5 (C2-3), 20.6 (C8-9). IR υ_{max} 3307, 2969, 1738, 1641, 1525, 1248. HRMS (ESI) found 303.1307 $C_{14}H_{20}N_2O_4Na^+$ $[M+Na]^+$ requires 303.1315.

***tert*-Butyl(4-(3-(3-(3-azidopropoxy)-4-(3-isobutyl-3-(3-isopropoxy-4-nitrobenzoyl)ureido)benzoyl)-1-(*tert*-butoxycarbonyl)-3-isobutylureido)-3-(3-(bis(*tert*-butoxycarbonyl)amino) propoxy)benzoyl)(isobutyl)carbamate (87)**

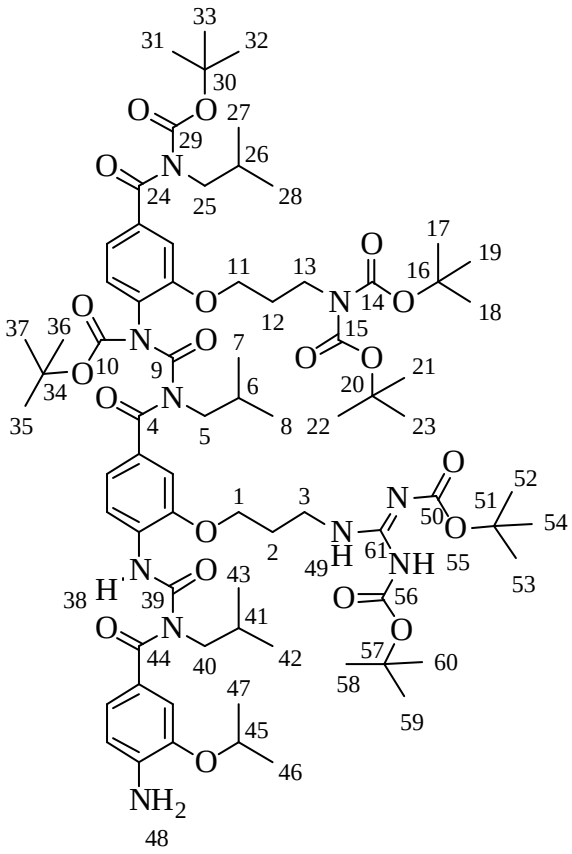


Following the general procedure for benzoylurea formation, amide **86** (52 mg, 0.19 mmol, 1.0 eq) was reacted with isocyanate **84** (187 mg, 0.19 mmol, 1.0 eq) and purified (2:1 petrol:ether) to yield the title compound as a yellow oil (143 mg, 0.11 mmol, 60 %): δ_H (400 MHz, $CDCl_3$) 11.75 (1H, s, H38), 8.43 (1H, d, *J* 8.4, Ar), 7.83 (1H, d, *J* 8.2, Ar), 7.31 (1H, dd, *J* 8.4; 1.6, Ar), 7.25 (1H, d, *J* 1.6, Ar), 7.16 (1H, d, *J* 1.1, Ar), 7.07 (1H, dd, *J* 8.3; 1.5, Ar), 7.05 (1H, d, *J* 1.6, Ar), 6.97 (1H, dd, *J* 8.0; 1.6, Ar), 6.56 (1H, d, *J* 8.0, Ar), 4.71 (1H, sept, *J* 6.1, H45), 4.14 (2H, t, *J* 5.7, H1 or H11), 3.97 (2H, t, *J* 7.1, H1 or H11), 3.71-3.55 (10H, m, H3, H5, H13, H25, and H40), 2.29-1.91 (7H, m, H2, H6, H12, H26, and H41), 1.48 (18H, s, H17-19 and H21-23), 1.40 (6H, d, *J*

6.0, H46-47), 1.28 (9H, s, H35-37), 1.19 (9H, s, H31-33), 0.96 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43), 0.93 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43), 0.81 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43). δ_C (100 MHz, $CDCl_3$) 173.5 (C4 or C24 or C44), 172.5 (C4 or C24 or C44), 171.5 (C4 or C24 or C44), 155.9 (C10), 154.5 (C14-15 or C29), 154.0 (C14-15 or C29), 152.6 (C9 or C39), 151.4 (C9 or C39), 151.3 (Ar), 151.1 (Ar), 147.5 (Ar), 142.1 (Ar), 140.8 (Ar), 139.1 (Ar), 131.4 (Ar), 131.3 (Ar), 130.7 (Ar), 127.5 (Ar), 125.9 (Ar), 122.3 (Ar), 119.7 (Ar), 118.6 (Ar), 118.4 (Ar), 114.8 (Ar), 111.6 (Ar), 111.5 (Ar), 83.6 (C30 or C34), 83.3 (C30 or C34), 82.5 (C16 and C20), 73.3 (C45), 67.1 (C1 or C11), 65.5 (C1 or C11), 55.6 (C5 or C25 or C40), 54.3 (C5 or C25 or C40), 52.9 (C5 or C25 or C40), 48.1 (C3 or C13), 43.4 (C3 or C13), 29.0 (C6 or C26 or C41), 28.8 (C6 or C26 or C41), 28.7 (C6 or C26 or C41), 28.4 (C2 or C12), 28.3 (C2 or C12), 28.2 (C17-19 and C21-23), 27.9 (C31-33), 27.6 (C35-37), 21.8 (C46-47), 20.5 (C7-8 or C27-28 or C42-43), 20.3 (C7-8 or C27-28 or C42-43), 20.3 (C7-8 or C27-28 or C42-43). IR υ_{max} 3220, 2964, 2100, 1715, 1665, 1533, 1367, 1122. HRMS (ESI) found 1311.6468 $C_{64}H_{92}N_{10}O_{18}Na^+$ $[M+Na]^+$ requires 1311.6483.

tert-Butyl(4-(3-(3-(4-amino-3-(3-(bis(_tert_-butoxycarbonyl)guanidinopropoxy)-4-(3-isobutyl-3-(3-isopropoxy-4-nitrobenzoyl)ureido)benzoyl)-1-(_tert_-butoxycarbonyl)-3-isobutylureido)-3-(3-(bis(_tert_-butoxycarbonyl)amino)propoxy)benzoyl)(isobutyl) carbamate (88)

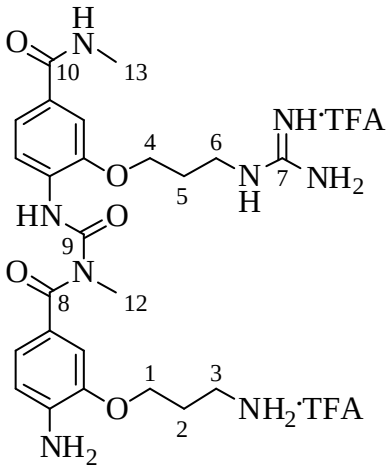
N,*N*'-diBoc-1*H*-pyrazole-1-carboxamidine (14 mg, 0.044 mmol, 1.2 eq) and palladium on activated carbon (47 mg, 5 % Pd by mass) were added to a stirred solution of benzoylurea **87** (47 mg, 0.036 mmol, 1.0 eq) in EtOAc (4.0 mL). The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen for 67 hours. The solution was filtered through Celite® with EtOAc (50 mL), concentrated, and purified *via* flash column chromatography (2:1 ether:petrol) to yield title compound as a colorless oil



(48 mg, 0.033 mmol, 89 %): δ_H (500 MHz, $CDCl_3$) 11.48 (1H, s, H55), 11.44 (1H, s, H38), 8.48 (1H, t, *J* 5.2, H49), 8.45 (1H, d, *J* 8.4, Ar), 7.38 (1H, dd, *J* 8.4; 1.7, Ar), 7.24 (1H, d, *J* 1.6, Ar), 7.07 (1H, d, *J* 1.6, Ar), 7.05 (1H, dd, *J* 8.1; 1.8, Ar), 7.00-6.97 (2H, m, Ar), 6.67 (1H, d, *J* 8.1, Ar), 6.63 (1H, d, *J* 8.0, Ar), 4.58 (1H, sept, *J* 6.0, H45), 4.14 (2H, br s, H48), 4.08 (2H, t, *J* 6.0, H11), 3.98 (2H, t, *J* 7.2, H1), 3.82 (2H, d, *J* 7.1, H5 or H25 or H40), 3.73-3.64 (6H, m, H3 and H13 and H5 or H25 or H40), 3.60 (2H, d, *J* 7.3, H5 or H25 or H40), 2.28-1.95 (7H, m, H2, H6, H12, H26, and H41), 1.50(18H, s, H17-19 and H21-23), 1.48 (9H, s, H52-54 or H58-60), 1.45 (9H, s, H52-54 or H58-60), 1.35 (6H, d, *J* 6.0, H46-47), 1.25 (9H, s, H35-37), 1.19 (9H, s, H31-33), 0.96 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43), 0.94 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43), 0.81 (6H, d, *J* 6.7, H7-8 or H27-28 or H42-43). δ_C (125 MHz, $CDCl_3$) 175.6 (C4 or C24 or C44), 172.6 (C4 or C24 or C44), 171.8 (C4 or C24 or C44), 163.6 (C61), 156.4 (C10), 156.0 (C14-15 or C29 or C50 or C56), 154.6 (C14-15 or C29 or C50 or C56), 154.1 (C14-15 or C29 or C50 or C56), 153.2 (C14-15 or C29 or C50 or C56), 152.8 (C9 or C39), 152.6 (C9 or C39), 151.1 (Ar), 147.5 (Ar), 144.3 (Ar), 141.2 (Ar), 139.0 (Ar), 132.2 (Ar), 130.9 (Ar), 130.4 (Ar), 127.5 (Ar), 124.4 (Ar), 122.7 (Ar), 122.3 (Ar), 119.9 (Ar), 118.0 (Ar), 113.8 (Ar), 113.4 (Ar), 111.8 (Ar), 111.6 (Ar), 83.5 (C16 and C20 or C30 or C34 or C51 or C57), 83.4 (C16 and C20 or C30 or C34 or C51 or C57), 83.2 (C16 and C20 or C30 or C34 or C51 or C57), 82.6 (C16 and C20 or C30 or C34 or C51 or C57), 79.5 (C16 and C20 or C30 or C34 or C51 or C57),

71.0 (C45), 67.1 (C1), 66.4 (C11), 55.6 (C5 or C25 or C40), 55.2 (C5 or C25 or C40), 52.9 (C5 or C25 or C40), 43.4 (C3), 37.9 (C13), 29.1 (C2 or C12), 28.9 (C2 or C12), 28.7 (C6 or C26 or C41), 28.4 (C52-54 or C58-60), 28.3 (C6 or C26 or C41), 28.3 (C17-19 and C21-23), 28.2 (C6 or C26 or C41), 28.2 (C52-54 or C58-60), 27.9 (C31-33), 27.6 (C35-37), 22.2 (C46-47), 20.5 (C7-8 or C27-28 or C42-43), 20.3 (C7-8 or C27-28 or C42-43), 20.1 (C7-8 or C27-28 or C42-43). IR υ_{max} 3338, 2976, 1731, 1667, 1532, 1368, 1329, 1133. HRMS (ESI) found 1497.8088 $C_{75}H_{114}N_{10}O_{20}Na^+$ $[M+Na]^+$ requires 1497.8103.

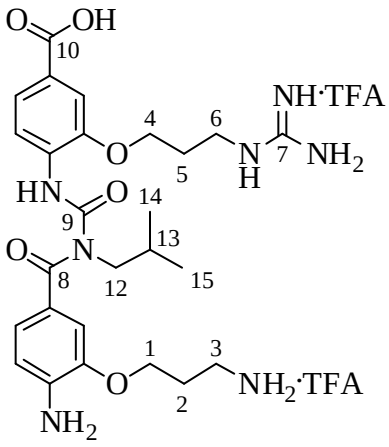
4-Amino-3-(3-aminopropoxy)-*N*-((2-(3-guanidinopropoxy)-4-(methylcarbamoyl)phenyl) carbamoyl)-*N*-methylbenzamide•2TFA (37)



Following the general procedure for *N*-boc deprotection, protected benzoylurea **76** (16 mg, 0.016 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (15 % MeCN to 60 % MeCN over 40 minutes) to yield the title compound (7 mg, 9.7 μmol) as an orange amorpohous solid: δ_H (500 MHz, D_2O) 7.96 (1H, d, *J* 8.4, Ar), 7.38 (1H, d, *J* 8.1, Ar), 7.29 (1H, dd, *J* 8.4; 1.8, Ar), 7.26 (1H, d, *J* 1.5, Ar), 7.21 (1H, br s, Ar), 7.17 (1H, dd, *J* 8.1; 1.6, Ar), 4.18 (2H, t, *J* 5.8, H1 or H4), 4.06 (2H, t, *J* 5.7, H1 or H4), 3.31 (2H, t, *J* 7.0, H3 or H6), 3.19-3.13 (5H, m, H3 or H6 and H12 or H13), 2.80 (3H, s, H12 or H13), 2.15 (2H, quint, *J* 7.3, H2 or H5), 2.00 (2H, quint, *J* 6.3, H2 or H5). δ_C (125 MHz, D_2O) 174.7 (C8), 169.8 (C10), 162.9 (q, *J* 35.6, $F_3C\mathit{C}O_2$), 156.7 (C7), 153.4 (C9), 150.7 (Ar), 148.0 (Ar), 135.0 (Ar), 129.7 (Ar), 129.5 (Ar), 124.4 (Ar), 123.2 (Ar), 120.2 (Ar), 120.1 (Ar), 119.6 (Ar), 116.3 (q, *J* 291.5, $F_3\mathit{C}CO_2$), 111.1 (Ar), 110.2 (Ar), 66.2 (C1 or C4), 65.9 (C1 or C4), 38.1 (C3 or C6), 36.9 (C3 or C6), 35.4 (C12 or C13), 27.7 (C2 or C5), 26.4 (C2 or C5), 26.3 (C12 or C13).

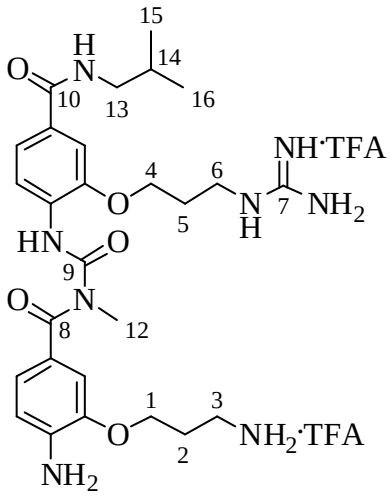
IR υ_{max} 3361, 3186, 2956, 1677, 1632, 1531, 1202, 1136. HRMS (ESI) found 515.2722 $C_{24}H_{35}N_8O_5^+$ $[M+H]^+$ requires 515.2725.

4-(3-(4-Amino-3-(3-aminopropoxy)benzoyl)-3-isobutylureido)-3-(3-guanidinopropoxy) benzoic acid•2TFA (38)



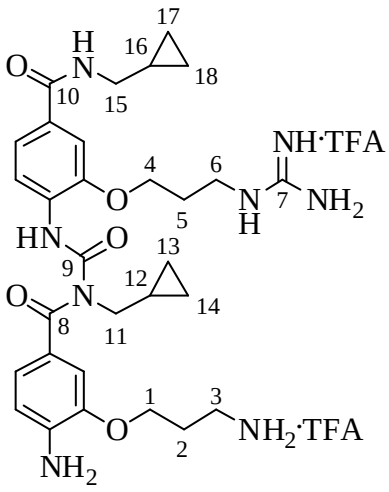
Following the general procedure for *N*-Boc deprotection, protected benzoylurea **77** (51 mg, 0.051 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (18 mg, 0.023 mmol) as a viscous orange oil: δ_H (500 MHz, D_2O) 7.73 (1H, d, *J* 8.4, Ar), 7.45-7.39 (2H, m, Ar), 7.28 (1H, d, *J* 5.0, Ar), 7.22 (1H, s, Ar), 7.18 (1H, d, *J* 8.1, Ar), 4.16 (2H, t, *J* 5.7, H1 or H4), 3.92 (2H, d, *J* 4.5, H1 or H4), 3.56 (2H, d, *J* 7.1, H12), 3.26 (2H, t, *J* 7.1, H3 or H6), 3.14 (2H, t, *J* 7.6, H3 or H6), 2.14 (2H, quint, *J* 7.0, H2 or H5), 1.94 (2H, quint, *J* 6.2, H2 or H5), 1.79 (1H, sept, *J* 6.5, H13), 0.71 (6H, d, *J* 6.5, H14-15). δ_C (125 MHz, D_2O) 174.7 (C8), 169.4 (C10), 162.8 (q, *J* 35.6, F_3CCO_2), 156.7 (C7), 153.0 (C9), 151.1 (Ar), 147.8 (Ar), 135.9 (Ar), 130.8 (Ar), 125.6 (Ar), 123.8 (Ar), 123.2 (Ar), 123.0 (Ar), 120.3 (Ar), 119.4 (Ar), 116.3 (q, *J* 291.7, F_3CCO_2), 112.0 (Ar), 111.3 (Ar), 66.3 (C1 or C4), 65.7 (C1 or C4), 54.1 (C12), 38.2 (C3 or C6), 36.8 (C3 or C6), 28.1 (C13), 27.7 (C2 or C5), 26.4 (C2 or C5), 19.1 (C14-15). IR υ_{max} 3370, 3191, 2964, 1676, 1632, 1540, 1190, 1137. HRMS (ESI) found 544.2878 $C_{26}H_{38}N_7O_6^+$ $[M+H]^+$ requires 544.2878.

4-Amino-3-(3-aminopropoxy)-*N*-((2-(3-guanidinopropoxy)-4-isobutylcarbamoyl) phenyl)carbamoyl)-*N*-methylbenzamide•2TFA (39)



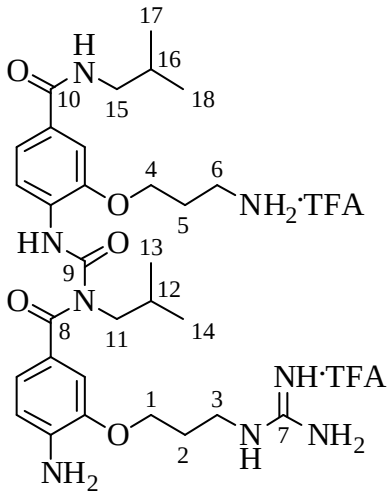
Following the general procedure for *N*-Boc deprotection, protected benzoylurea **78** (58 mg, 0.055 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (25 mg, 0.032 mmol) as an amorphous white solid: δ_H (500 MHz, D_2O) 7.94 (1H, d, *J* 8.5, Ar), 7.44 (1H, d, *J* 8.1, Ar), 7.25-7.20 (2H, m, Ar), 7.15 (1H, d, *J* 1.5, Ar), 7.13 (1H, dd, *J* 8.1; 1.3, Ar), 4.20 (2H, t, *J* 5.7, H1 or H4), 3.94 (2H, t, *J* 5.6, H1 or H4), 3.26 (2H, t, *J* 7.3, H3 or H6), 3.14 (2H, t, *J* 7.6, H3 or H6), 3.06 (3H, s, H12), 3.02 (2H, d, *J* 7.1, H13), 2.16 (2H, quint, *J* 7.6, H2 or H5), 1.92 (2H, quint, *J* 6.7, H2 or H5), 1.71 (1H, sept, *J* 6.8, H14), 0.78 (6H, d, *J* 6.7, H15-16). δ_C (125 MHz, D_2O) 174.6 (C8), 168.8 (C10), 162.2 (q, *J* 35.6, $F_3C\mathit{C}O_2$), 156.6 (C7), 152.6 (C9), 151.3 (Ar), 147.5 (Ar), 136.0 (Ar), 129.8 (Ar), 129.2 (Ar), 124.1 (Ar), 122.7 (Ar), 120.0 (Ar), 119.8 (Ar), 118.9 (Ar), 116.2 (q, *J* 291.7, $F_3\mathit{C}CO_2$), 111.0 (Ar), 109.7 (Ar), 66.3 (C1 or C4), 65.8 (C1 or C4), 47.3 (C13), 38.2 (C3 or C6), 36.9 (C3 or C6), 35.5 (C12), 28.1 (C14), 27.8 (C2 or C5), 26.4 (C2 or C5), 19.3 (C15-16). IR υ_{max} 3353, 3189, 3079, 2963, 1674, 1630, 1528, 1201, 1136. HRMS (ESI) found 557.3198 $C_{27}H_{41}N_8O_5^+$ $[M+H]^+$ requires 557.3194.

4-Amino-3-(3-aminopropoxy)-*N*-(cyclopropylmethyl)-*N*-((4-((cyclopropylmethyl) carbamoyl)-2-(3-guanidinopropoxy)phenyl)carbamoyl)benzamide•2TFA (40)



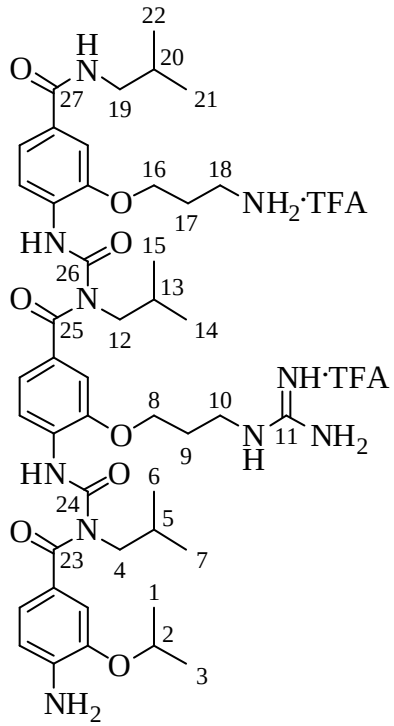
Following the general procedure for *N*-Boc deprotection, protected benzoylurea **74** (32 mg, 0.029 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (15 mg, 0.018 mmol) as a viscous orange oil: δ_H (500 MHz, D_2O) 7.80 (1H, d, *J* 8.3, Ar), 7.48 (1H, d, *J* 8.1, Ar), 7.34 (1H, dd, *J* 8.4; 1.7, Ar), 7.32 (1H, br s, Ar), 7.31-7.27 (2H, m, Ar), 4.23 (2H, t, *J* 5.7, H1 or H4), 4.09 (2H, t, *J* 5.7, H1 or H4), 3.72 (2H, d, *J* 7.3, H11 or H15), 3.36 (2H, t, *J* 7.0, H3 or H6), 3.24-3.19 (4H, m, H3 or H6 and H11 or H15), 2.25-2.17 (2H, m, H2 or H5), 2.08-2.01 (2H, m, H2 or H5), 1.14-1.00 (2H, m, H12 and H16), 0.54-0.46 (4H, m, H13-14 or H17-18), 0.23 (2H, dd, *J* 10.0; 4.9, H13-14 or H17-18), 0.16 (2H, dd, *J* 10.0; 5.0, H13-14 or H17-18). δ_C (125 MHz, D_2O) 173.9 (C8), 169.3 (C10), 162.9 (q, *J* 35.6, $F_3C\mathit{C}O_2$), 156.7 (C7), 154.0 (C9), 151.0 (Ar), 148.7 (Ar), 135.9 (Ar), 130.6 (Ar), 129.0 (Ar), 123.7 (Ar), 123.3 (Ar), 120.6 (Ar), 120.1 (Ar), 120.0 (Ar), 116.3 (q, *J* 291.6, $F_3\mathit{C}CO_2$), 111.2 (Ar), 110.5 (Ar), 66.2 (C1 or C4), 65.8 (C1 or C4), 51.6 (C11 or C15), 44.6 (C11 or C15), 38.1 (C3 or C6), 36.8 (C3 or C6), 27.7 (C2 or C5), 26.4 (C2 or C5), 10.1 (C12 or C16), 9.8 (C12 or C16), 3.2 (C13-14 or C17-18), 2.6 (C13-14 or C17-18). IR υ_{max} 3358, 3187, 2963, 1676, 1631, 1525, 1200, 1136. HRMS (ESI) found 595.3360 $C_{30}H_{43}N_8O_5^+$ $[M+H]^+$ requires 595.3351.

4-Amino-*N*-((2-(3-aminopropoxy)-4-(isobutylcarbamoyl)phenyl)carbamoyl)-3-(3-guanidinopropoxy)-*N*-isobutylbenzamide•2TFA (41)



Following the general procedure for *N*-Boc deprotection, protected benzoylurea **75** (24 mg, 0.022 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (17 mg, 0.021 mmol) as a viscous orange oil: δ_H (500 MHz, D_2O) 7.77 (1H, d, *J* 8.2, Ar), 7.48 (1H, d, *J* 8.2, Ar), 7.34 (1H, dd, *J* 8.5; 1.9, Ar), 7.32-7.29 (2H, m, Ar), 7.28 (1H, br s, Ar), 4.17-4.11 (4H, m, H1 and H4), 3.69 (2H, d, *J* 8.3, H11 or H15), 3.34 (2H, t, *J* 7.7, H3 or H6), 3.20 (2H, t, *J* 7.4, H3 or H6), 3.15 (2H, d, *J* 7.3, H11 or H15), 2.18 (2H, quint, *J* 6.8, H2 or H5), 2.07 (2H, quint, *J* 6.0, H2 or H5), 1.96 (1H, sept, *J* 6.7, H12 or H16), 1.86 (1H, sept, *J* 6.8, H12 or H16), 0.90 (6H, d, *J* 7.3, H13-14 or H17-18), 0.85 (6H, d, *J* 6.6, H13-14 or H17-18). δ_C (125 MHz, D_2O) 174.0 (C8), 169.5 (C10), 162.8 (q, *J* 35.6, $F_3C\mathit{C}O_2$), 156.8 (C7), 154.3 (C9), 151.1 (Ar), 148.7 (Ar), 136.0 (Ar), 131.0 (Ar), 128.9 (Ar), 123.8 (Ar), 123.3 (Ar), 120.8 (Ar), 120.4 (Ar), 120.0 (Ar), 116.3 (q, *J* 291.6, $F_3\mathit{C}CO_2$), 111.3 (Ar), 110.6 (Ar), 66.1 (C1 or C4), 65.7 (C1 or C4), 54.2 (C11 or C15), 47.3 (C11 or C15), 37.8 (C3 or C6), 36.8 (C3 or C6), 28.0 (C12 and C16), 27.5 (C2 or C5), 26.5 (C2 or C5), 19.3 (C13-14 or C17-18), 19.1 (C13-14 or C17-18). IR υ_{max} 3351, 3190, 2964, 1680, 1587, 1524, 1201, 1137. HRMS (ESI) found 599.3671 $C_{30}H_{47}N_8O_5^+$ $[M+H]^+$ requires 599.3664.

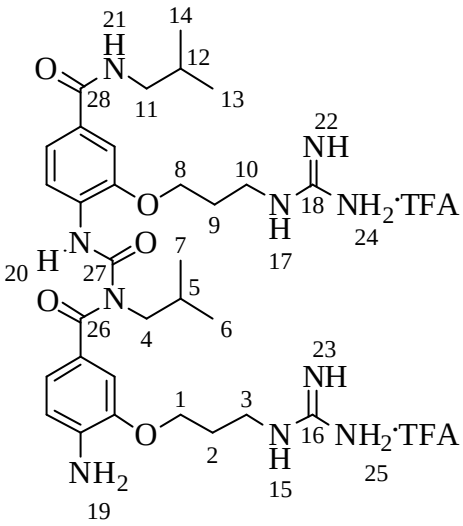
4-Amino-*N*-((4-(((2-(3-aminopropoxy)-4-(isobutylcarbamoyl)phenyl)carbamoyl) (isobutyl) carbamoyl)-2-(3-guanidinopropoxy)phenyl)carbamoyl)-*N*-isobutyl-3-isopropoxybenzamide (42)



Following the general procedure for *N*-Boc deprotection, protected benzoylurea **88** (46 mg, 0.031 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (17 mg, 0.015 mmol) as an orange oil: δ_H (500 MHz, MeOD) 8.42 (1H, d, *J* 8.9, Ar), 8.36 (1H, d, *J* 8.4, Ar), 7.54 (1H, d, *J* 1.7, Ar), 7.52 (1H, dd, *J* 8.4; 1.8, Ar), 7.28-7.24 (2H, m, Ar), 7.11-7.06 (2H, m, Ar), 6.84 (1H, d, *J* 8.1, Ar), 4.65 (1H, sept, *J* 6.1, H2), 4.30 (2H, t, *J* 5.7, H8 or H16), 4.22 (2H, t, *J* 5.6, H8 or H16), 3.87 (2H, d, *J* 7.2, H4 or H12 or H19), 3.82 (2H, d, *J* 7.2, H4 or H12 or H19), 3.59 (2H, t, *J* 7.4, H10 or H18), 3.36 (2H, t, *J* 7.4, H10 or H18), 3.23 (2H, d, *J* 7.1, H4 or H12 or H19), 2.30 (2H, quint, *J* 7.5, H9 or H17), 2.19 (2H, quint, *J* 7.6, H9 or H17), 2.02-1.90 (3H, m, H5 and H13 and H20), 1.38 (6H, d, *J* 6.1, H1 and H3), 0.99 (6H, d, *J* 6.7, H6-7 or H14-15 or H21-22), 0.83 (12H, d, *J* 6.7, two of H6-7 and H14-15 and H21-22). δ_C (125 MHz, MeOD) 177.9 (C23 or C25), 176.9 (C23 or C25), 169.6 (C27), 162.8 (q, *J* 34.3, F_3CCO_2), 158.7 (C11), 154.1 (C24 or C26), 153.6 (C24 or C26), 148.9 (Ar), 148.9 (Ar), 145.8 (Ar), 143.3 (Ar), 132.0 (Ar), 132.0 (Ar), 131.9 (Ar), 131.1 (Ar), 124.8 (Ar), 123.7 (Ar), 122.2 (Ar), 121.4 (Ar), 120.3 (Ar), 120.2 (Ar), 117.2 (Ar), 115.2 (Ar), 114.6 (Ar), 111.4 (Ar), 72.3 (C2), 67.3 (C8 or C16), 66.8 (C8 or C16), 58.4 (C4 or C12 or C19), 56.2 (C4 or C12 or C19), 55.6 (C4 or C12 or C19), 39.6 (C10 or C18), 38.2 (C10 or C18), 29.9 (C5 or C13 or C20), 29.8 (C5 or C13 or C20), 29.8 (C5 or C13 or C20), 29.7 (C9 or C17), 28.5 (C9 or

C17), 22.3 (C1 and C3), 20.6 (C6-7 or C14-15 or C21-22), 20.3 (C6-7 or C14-15 or C21-22), 20.2 (C6-7 or C14-15 or C21-22). IR υ_{max} 3362, 3196, 2962, 1679, 1630, 1587, 1201, 1190, 1136. HRMS (ESI) found 875.5109 $C_{45}H_{67}N_{10}O_8^+$ $[M+H]^+$ requires 875.5138.

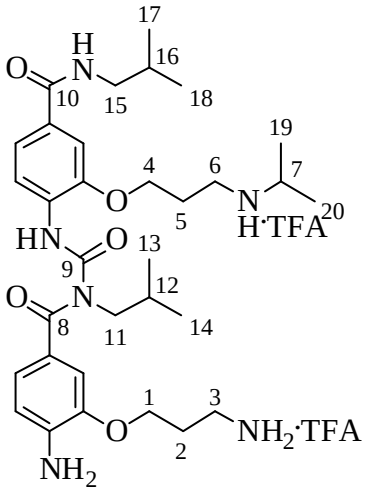
4-amino-3-(3-guanidinopropoxy)-*N*-((2-(3-guanidinopropoxy)-4-(isobutylcarbamoyl)phenyl) carbamoyl)-*N*-isobutylbenzamide • 2 TFA (43)



Following the general procedure for *N*-Boc deprotection, benzoylurea **81** (94 mg, 0.082 mmol) was stirred for 90 minutes and the resulting residue purified *via* semi-preparative HPLC (35 % MeCN to 80 % MeCN over 40 minutes) to yield the title compound as a colorless wax: δ_H (500 MHz, *d3*-MeCN) 11.33 (1H, s, H20), 8.33 (1H, d, *J* 8.3, Ar), 7.99 (1H, t, *J* 5.0, H15 or H17), 7.72 (1H, t, *J* 5.1, H15 or H17), 7.42 (1H, dd, *J* 8.5; 1.8, Ar), 7.40 (1H, d, *J* 1.6, Ar), 7.08 (1H, t, *J* 5.6, H21), 7.03 (1H, dd, *J* 8.3; 1.7, Ar), 7.00 (1H, d, *J* 1.7, Ar), 6.91 (8H, br s, H19 and H22-25), 6.71 (1H, d, *J* 8.1, Ar), 4.19 (2H, t, *J* 5.7, H1 or H8), 4.06 (2H, t, *J* 5.7, H1 or H8), 3.80 (2H, d, *J* 7.3, H4), 3.50 (2H, q, *J* 6.3, H3 or H10), 3.35 (2H, q, *J* 6.5, H3 or H10), 3.16 (2H, t, *J* 6.6, H11), 0.93 (6H, d, *J* 6.6, H13-14), 0.77 (6H, d, *J* 6.9, H6-7). δ_H (500 MHz, D_2O) 7.68 (1H, d, *J* 8.5, Ar), 7.38 (1H, d, *J* 8.2, Ar), 7.35 (1H, dd, *J* 8.3; 1.9, Ar), 7.33-7.29 (2H, m, Ar), 7.24 (1H, s, Ar), 4.14-4.09 (4H, m, H1 and H8), 3.72 (2H, d, *J* 7.3, H4), 3.38-3.31 (4H, m, H3 and H10), 3.18 (2H, d, *J* 6.9, H11), 2.09-1.97 (5H, m, H2 and H9 and H12), 1.88 (1H, sept, *J* 6.6, H5), 0.92 (6H, d, *J* 6.9, H13-14), 0.89 (6H, d, *J* 6.9, H6-7). δ_C (125 MHz, D_2O) 173.8 (C26), 169.6 (C28), 162.9 (q, CF_3CO_2, J_{CF} 35 Hz),156.8 (C27 or C16 and C18), 156.7 (C27 or C16 and C18), 150.1 (Ar), 149.3 (Ar), 133.9 (Ar), 131.4 (Ar), 128.7 (Ar), 122.2

(Ar), 122.2 (Ar), 121.3 (Ar), 121.0 (Ar), 119.9 (Ar), 116.3 (q, *J* 292, *C*F_3CO_2), 111.5 (Ar), 110.7 (Ar), 66.1 (C1 or C8), 65.8 (C1 or C8), 54.2 (C4), 47.3 (C11), 38.1 (C3 or C10), 37.9 (C3 or C10), 28.0 (C5 or C12), 28.0 (C5 or C12), 27.6 (C2 or C9), 27.5 (C2 or C9), 19.3 (C6-7 or C13-14), 19.2 (C6-7 or C13-14). IR υ_{max} 3350, 3194, 2962, 1675, 1628, 1522, 1202. HRMS (ESI) found 641.3865 $C_{31}H_{49}N_{10}O_5^+$ $[M+H]^+$ requires 641.3882.

4-Amino-3-(3-aminopropoxy)-*N*-isobutyl-*N*-((4-(isobutylcarbamoyl)-2-(3-isopropylamino)propoxy)phenyl)carbamoyl)benzamide•2TFA (79)



Following the general procedure for *N*-Boc deprotection, protected benzoylurea **73** (31 mg, 0.034 mmol) was stirred for 90 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 95 % MeCN over 40 minutes) to yield the title compound (13 mg, 0.016 mmol) as a viscous orange oil: δ_H (500 MHz, D_2O) 7.77 (1H, d, *J* 8.2, Ar), 7.36 (1H, d, *J* 8.2, Ar), 7.28 (1H, dd, *J* 8.4; 1.8, Ar), 7.23 (1H, d, *J* 1.7, Ar), 7.22-7.19 (2H, m, Ar), 4.14 (2H, t, *J* 5.7, H1 or H4), 4.10 (2H, t, *J* 5.6, H1 or H4), 3.62 (2H, d, *J* 7.3, H11 or H15), 3.24 (1H, sept, *J* 6.6, H7), 3.20-3.11 (4H, m, H3 and H6), 3.08 (2H, d, *J* 7.1, H11 or H15), 2.17-2.07 (4H, m, H2 and H5), 1.86 (1H, sept, *J* 6.8, H12 or H16), 1.79 (1H, sept, *J* 6.8, H12 or H16), 1.14 (6H, d, *J* 6.6, H19-20), 0.83 (6H, d, *J* 6.7, H13-14 or H17-18), 0.75 (6H, d, *J* 6.7, H13-14 or H17-18). δ_C (125 MHz, D_2O) 174.7 (C8), 169.6 (C10), 162.9 (q, *J* 35.6, F_3CCO_2), 154.0 (C9), 153.9 (Ar), 150.7 (Ar), 148.5 (Ar), 135.1 (Ar), 130.7 (Ar), 129.1 (Ar), 124.8 (Ar), 123.0 (Ar), 120.6 (Ar), 120.1 (Ar), 116.3 (q, *J* 291.7, F_3CCO_2), 111.4 (Ar), 110.5 (Ar), 66.2 (C1 or C4), 65.6 (C1 or C4), 54.3 (C11 or C15), 50.8 (C7), 47.3 (C11 or C15), 41.9 (C3 or C6), 36.9 (C3 or C6), 28.1 (C12 or C16), 28.0 (C12 or C16), 26.4 (C2 or C5), 25.7

(C2 or C5), 19.3 (C13-14 or C17-18), 19.1 (C13-14 or C17-18), 18.1 (C19-20). IR υ_{max} 3352, 3201, 2962, 1678, 1630, 1524, 1202, 1135. HRMS (ESI) found 599.3899 $C_{32}H_{51}N_6O_5^+$ $[M+H]^+$ requires 599.3915.

4. Benzamide-based 3_{10} peptidomimetics

4.1. Introduction

The 3_{10} helix is a compact, steeply-rising fold adopted by polypeptides and proteins. It is a right-handed helix consisting of three residues per turn, rising 6 Å per turn.[1] Hydrogen bonding is between the backbones of the *i* and *i*+3 residues, and due to a 3-fold axis of symmetry, the sidechains of successive turns are projected in the same directions (Figure 4-1). While this helix maintains optimal intramolecular hydrogen bond lengths, it does so by twisting the amide C-N bond, to an extent that the structure was thought impossible by Pauling in his initial study of protein helices.[2] It is now known that the 3_{10} secondary structure is in fact common, making up 15 % - 20 % of helices in the PDB.[3] On average 3_{10} helices are quite short, three to five residues, compared to ten to twelve residues for α-helices, although longer stretches of the 3_{10} structure are known.[4] Further analysis suggests that the 3_{10} helix structure in proteins deviates from ideality, with 3_{10} stretches containing nine residues or more uniformly lacking long-range order.[5]

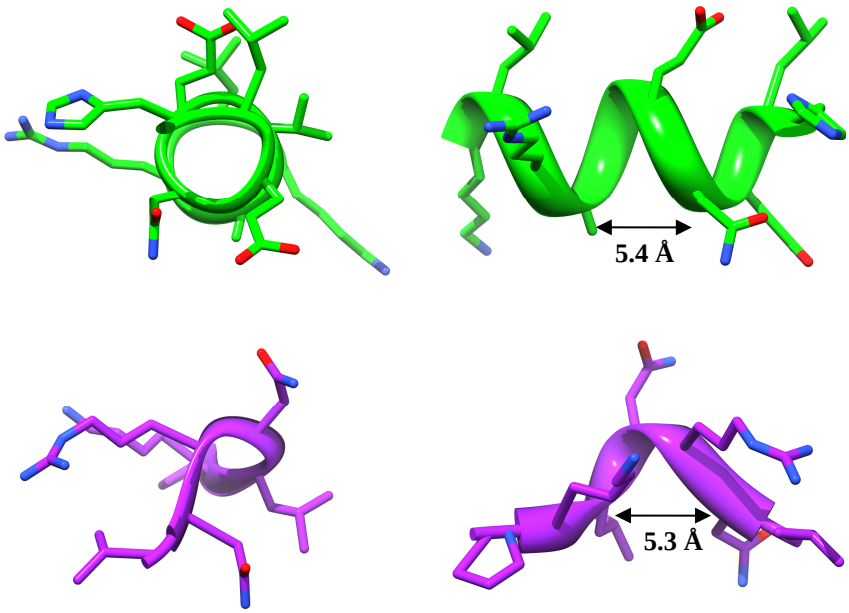


Figure 4-1: Comparison of an α-helix (green) and four residues of the 3_{10} helix of Gab2b (purple). The 3_{10} structure has a 3-fold axis of symmetry and is very narrow (bottom left) when compared to the unsymmetrical α-helix (top left). The Gab2b peptide spans 5.3 Å, which deviates from the ideal 6.0 Å for a 3_{10} helix.

Mimicry of the 3_{10} structure began with the observation that α-α-disubstituted amino acids show a high degree of helicity resulting from their constrained backbone conformation. Sequences rich in 2-aminoisobutyric acid (α-methylalanine) are rich in α-helix or 3_{10} helix conformation, depending on the precise sequence.[6,7] Synthetic α-α-disubstituted amino acids have been used in a variety of "stapling" approaches, including those *via* lactam bridges,[8] ring-closing metathesis,[9,10] and click chemistry,[11] which stabilize 3_{10} helical character. Recent work on "universal peptidomimetics" has included the 3_{10} helix among the mimicked structures, and has resulted in the synthesis of several classes of molecules.[12,13] Unfortunately, the conformational flexibility of these molecules means they suffer a great entropic penalty upon binding,[14] and so far their biological utility has not been demonstrated. One interesting conclusion from this work is that scaffolds previously designed for α-helix mimicry (e.g. the terphenyl of Hamilton *et al.*[15]) might in fact be more effective at matching a 3_{10} structure.[13]

4.2. Peptidomimetic design

While the relatively compact structure of the 3_{10} helix is important in allowing peptides to form optimal H-bonds, a peptidomimetic with a wholly synthetic backbone need only match the distance between adjacent sidechains and their projection along the same face of the scaffold. The Gab2b peptide, which has a binding affinity 3.2 μM for Grb2 SH3C (by ITC), adopts a four reside 3_{10} helix (Arg-Asn-Leu-Lys) containing the key Arg and Lys residues required for binding to Grb2 SH3C. None of the other amino acids in this helix make significant contributions to binding energy,[16] so their inclusion into a peptidomimetic scaffold may not be necessary.

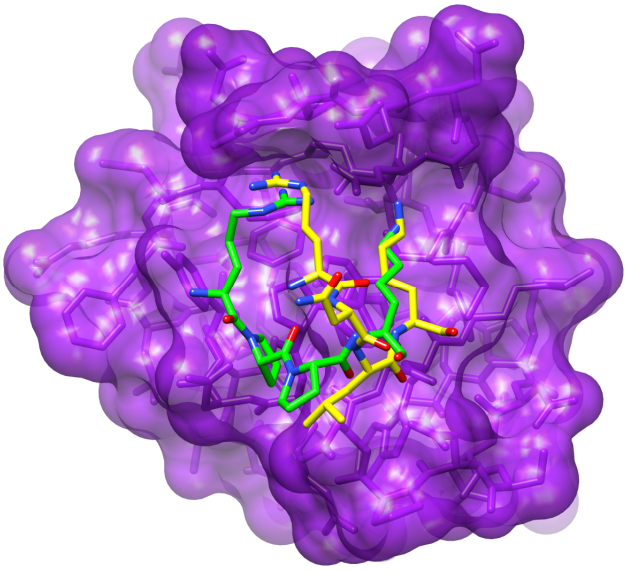


Figure 4-2: Comparison of the key residues of the Gab2a peptide (green; 2WVF) and the Gab2b peptide (yellow; 2W0Z) binding to the surface of Grb2 SH3C (purple). Displayed are the four-residue sequences of each peptide which contain the RxxK sequence. Due to the much greater length of a PPII helix relative to the 3_{10} helix, the two peptides display the cationic residues in different conformations to maintain optimal H-bond distances to the protein.

The shorter length of the 3_{10} helix of Gab2b relative to the PPII of Gab2a results in a different projection of the cationic residues into the binding groove of Grb2 SH3C (Figure 4-2). The protein structure distorts slightly to accommodate hydrogen bonds to the same key residues (*vide supra* Chapter 1.3), but the overall fold of Grb2 SH3C is not significantly changed when the peptides are bound. Nonetheless, the two peptide structures underscore the need for different peptidomimetic scaffolds for mimicking Gab2a and Gab2b.

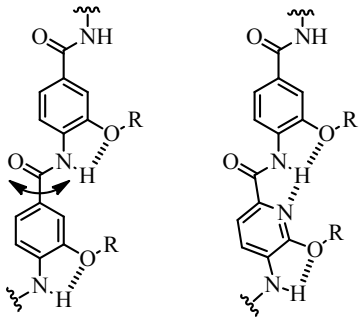


Figure 4-3: Comparison of the oligobenzamide scaffold (left) and the oligopyridylamide scaffold (right). The benzamide scaffold is less rigid as there is relatively unconstrained rotation around the Ar-C bond, whereas the conformation of the oligopyridylamide is largely fixed by a three-centre hydrogen bond. As a result of this constraint, however, the oligopyridylamide scaffold is also more curved than the oligobenzamide.[17]

Computational modelling led to consideration of the benzamide scaffold, which has been used extensively for peptidomimetics.[18] This scaffold is synthesized through sequential amide coupling reactions, and its properties such as curvature[18] and conformational flexibility[19] are fixed by intramolecular hydrogen bonding (Figure 4-3). For instance, the three-centre hydrogen bonds of the oligopyridylamide scaffold create a rigid structure projecting sidechains along one face, but at the cost of an overall curved backbone. In contrast, oligobenzamides are less conformationally constrained, and are correspondingly less curved.[17] This scaffold nonetheless shows a strong preference for forming the intramolecular hydrogen bond, even in high dielectric media such as water.[20] Synthesis of the benzamide scaffold is relatively simple,[21] and extensive work by multiple groups has allowed the incorporation of numerous sidechains,[22–24] so this scaffold was selected as a potential mimic of Gab2b.

Solid-state, NMR, and computational data all indicate that the benzamide conformation in which sidechains are projected along the same face is conformationally-accessible.[17,19,23] In this conformation, the distance between adjacent sidechains on the scaffold is 5.74 Å, which compares favourably to the 5.32 Å of the Gab2b peptide. The peptide forms only a single turn of a 3_{10} helix, so molecules incorporating two aromatic rings and two sidechains were chosen as the target of synthesis (Figure 4-4). Designed mimics investigate the role of hydrogen bonding by incorporating two Lys sidechains (**1**), mixed Arg and Lys sidechains (**2** and **3**), and two Arg sidechains (**4**). The Arg sidechains of **2** and **3** differ in length by one carbon: counting the ether oxygen as part of the chain, the shorter **2** correctly matches the length of Arg. Counting only carbon atoms, however, **3** is the appropriate length. Additionally, the role of the next key amino acid in the sequence, Pro, was studied by including cyclopentane (**5**) or Pro (**6**) as an additional residue. Overlaying a thermodynamically-accessible conformation of the proposed mimetic **3** (calculated at the

PM3 semi-empirical level of theory)[25] with the crystal structure of the bound Gab2b peptide gives an RMSD of the C_α and C_β atoms of both structures of 0.73 Å (for four points).

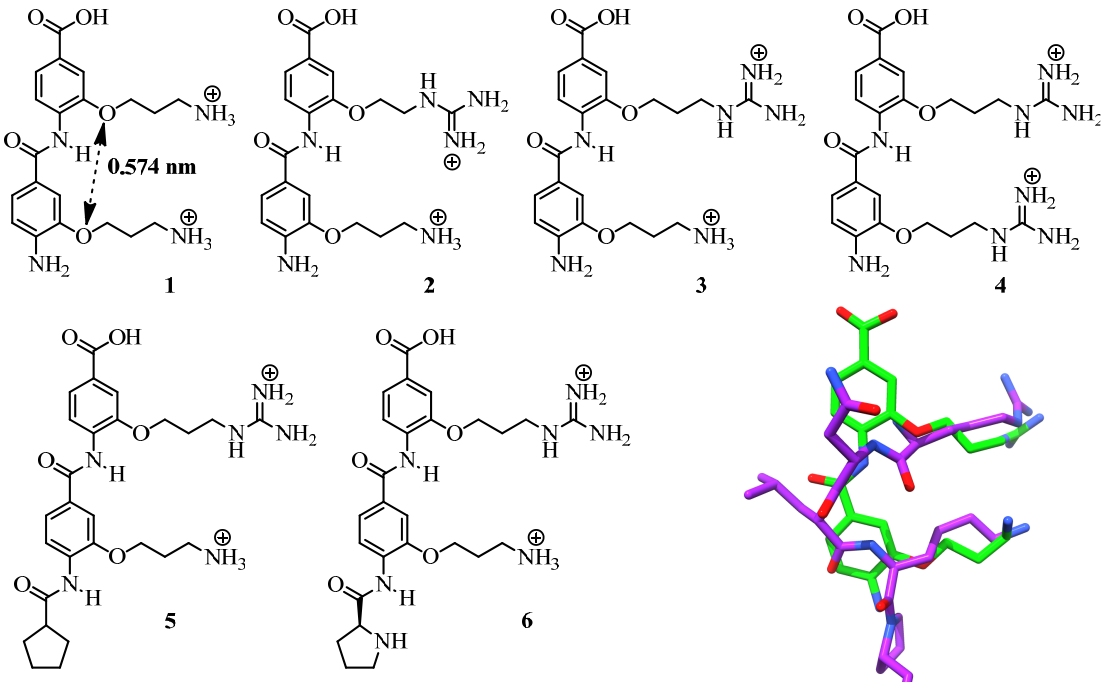
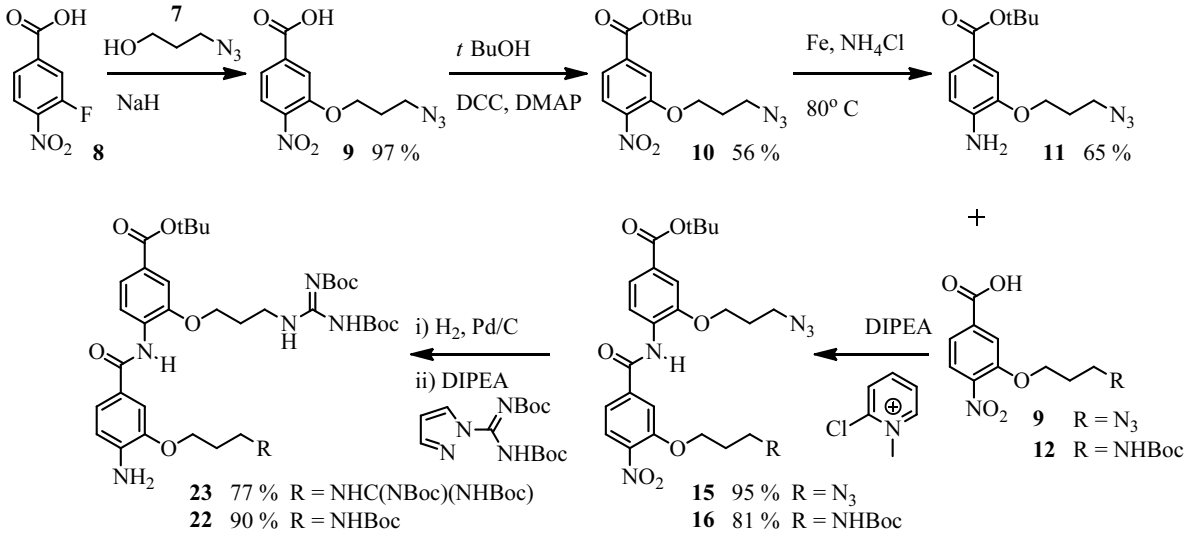


Figure 4-4: Structures of proposed mimics **1**-**6**. Bottom right: overlay of proposed mimic **3** (green) with the Gab2b peptide 3_{10} helix (purple).

4.3. Synthesis

The key amide bond in the scaffold is formed from a relatively electronically deactivated aniline. As a result, robust coupling conditions are required to form the benzamide. Early synthesis of this scaffold was not possible with two alkoxy sidechains,[21] but recent work has found that 2-chloro-1-methyl pyridinium iodide (Mukaiyama's reagent) allows facile synthesis of fully-functionalized benzamides.[23,26] Unlike the benzoylurea coupling described in Chapter 3, for which all labile protons required protection, in this case the coupling uses only weak base, so mildly acidic (pKa > ca. 10) sidechains are tolerated. For the planned Arg/Lys mimics, orthogonal protecting groups are required. Building from the benzoylurea synthesis (Chapter 3), an azide was selected as the first of these, while a single Boc group was sufficient to protect the other sidechain.

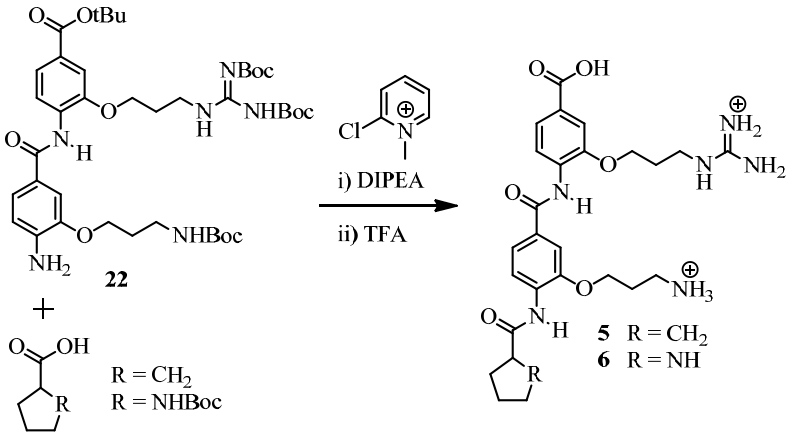
Synthesis began with nucleophilic aromatic substitution of azido-alcohol **7** with 3-fluoro-4-nitro-benzoic acid **8** (Scheme 4-1). The azide product **9** was the bottom fragment required for the eventual Lys/Lys and Arg/Arg mimics **1** and **4**. For the top fragment, the acid was then protected as the *tert*-butyl ester **10**. Reduction of the nitro group to the aniline furnished the completed fragment **11**. The carboxylic acid of the bottom fragment was then activated using Mukaiyama's reagent and DIPEA, and the amide bond formed with the aniline of the top fragment. Coupling azido-benzoic acid **9** to azido-aniline **11** yielded di-azide **15**, which was the precursor to the Lys/Lys and Arg/Arg mimics **1** and **4**. The first of these was formed *via* hydrogenation of the azide to the primary amine **18** followed by deprotection of the *tert*-butyl ester with TFA. Alternatively, the hydrogenated product could be reacted with *N*,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine to yield the protected di-Arg molecule **23**, which upon global deprotection with TFA became mimic **4**.



Scheme 4-1: Synthesis of benzamide mimics. Deprotection of **15** (with hydrogen and TFA) yields mimic **1**, while deprotection of **22** or **23** with TFA yields mimics **3** and **4**, respectively.

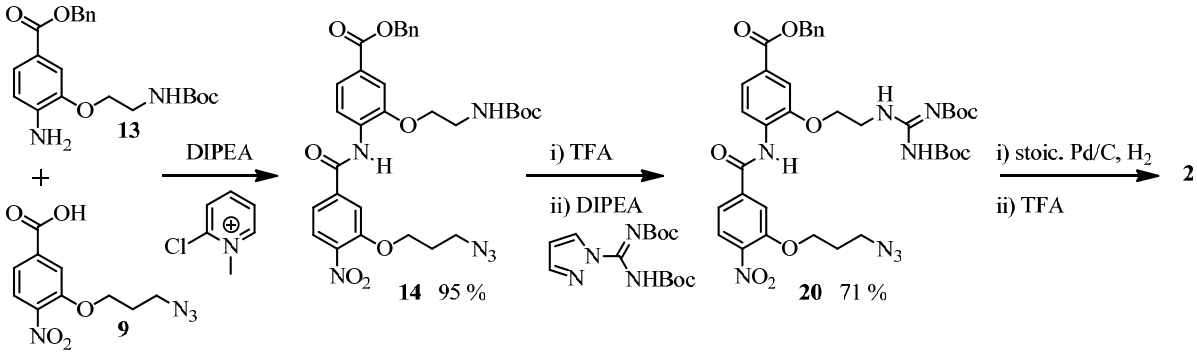
An analogous synthetic route, this time coupling the Boc-protected amine **12** (previously synthesized in the Hamilton group) to azide **11** yielded the orthogonally-protected **16**, which was the precursor to Arg/Lys mimics **3**, **5**, and **6** (Scheme 4-1). Synthesis of all three proceeded *via* hydrogenation of the azide to primary amine **19**, which

then furnished the protected Arg sidechain of **22**. Global deprotection with TFA yielded mimic **3**.



Scheme 4-2: Synthesis of the trifunctionalized Arg/Lys/Pro mimics.

Alternatively, **22** could be further functionalized to explore the role of additional hydrophobic groups (Scheme 4-2). Although not part of the 3_{10} helix, the next amino acid in the Gab2b peptide is Pro. Thus *N*-Boc protected L-proline was coupled to the benzamide using Mukaiyama's reagent, which, following TFA deprotection, formed mimic **6**. Cyclopentylcarboxylic acid, a potential isostere of proline, was separately coupled and deprotected with TFA to yield mimic **5.**



Scheme 4-3: Synthesis of the shorter Arg sidechain of mimic **2**.

The synthesis of the shorter Arg sidechain of **2** required more-complicated protecting group manipulation (Scheme 4-3). Synthesis of an ethoxy azide one carbon shorter than **7** was deemed unnecessarily hazardous, so instead Boc protection was pursued for the

eventual Arg sidechain. This required an acid-stable protecting group for the carboxylic acid of that fragment, so molecule **13** (previously synthesized in the Hamilton group) was used as the top fragment and **9** as the bottom half. These were coupled using Mukaiyama amide conditions, and the Boc group removed with TFA to amine **17**. The guanidine group was synthesized from this amine, and the final mimic synthesized by sequential deprotection (hydrogenation and TFA).

4.4. Biological evaluation of mimics

4.4.1. SPR

The mimics were initially screened for binding to Grb2 SH3C using SPR. Molecules were dissolved in running buffer containing 5 % DMSO and tested at 50 μM and 100 μM concentrations. Disappointingly, none showed strong binding, with a maximum response of 11.4 RU for mimic **1** (Table 4-1).

Table 4-1: SPR binding responses for putative 3_{10} mimics screened at 50 μM and 100 μM concentrations. None shows strong binding, with the maximum response for the Lys/Lys mimic **1** of 11.4 RU.

	1	**2**	**3**	**4**	**5**	**6**
50 μM	7.2	2.0	2.9	1.0	2.7	3.4
100 μM	11.4	4.4	6.9	1.3	5.5	7.6

In order to further characterize the binding of these mimics, a titration was conducted with mimics **1** and **3**, which showed relatively strong responses. Mimics were screened at concentrations from 6.25 μM to 200 μM for binding to both unlabelled Grb2 SH3C and ^{15}N-labeled protein. At low concentrations, the binding of mimic **3** was anomalous, with noise eclipsing signal (Figure 4-11). The affinity of the Lys/Lys mimic **1**, however, was calculated to be very similar for both surfaces (150 ± 60 μM and 150 ± 30 μM for the unlabelled and ^{15}N surfaces, respectively; Figure 4-5). The value of B_{max} (the maximum binding level, if the binding site were fully saturated) calculated from the fits, however, was

substantially different, with the unlabelled surface estimated to have two-fold greater B_{max}. While a higher concentration of one of the proteins would explain this discrepancy, in reality nearly identical amounts of each were bound. Potentially some of one of the proteins could be inactive, resulting in less total (but equal fractional) binding to the surface, but this cannot be directly measured once the proteins are attached to the SPR surface.

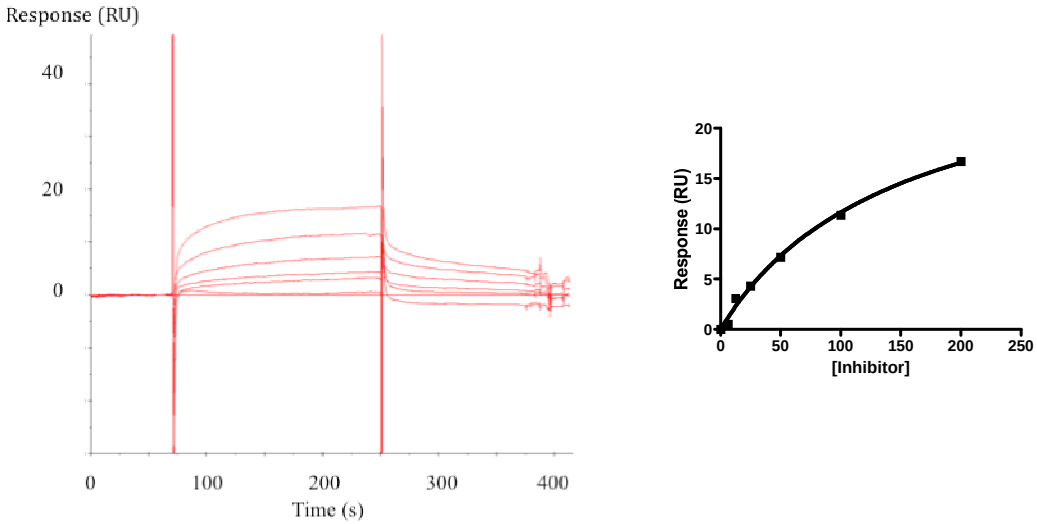


Figure 4-5: Left: sensograms of binding of mimic **1** to ^{15}N-Grb2 SH3C. Mimic concentrations are a two-fold dilution series from 200 μM (highest response) to 6.25 μM (lowest non-zero response). Right: fit of binding response vs. mimic concentration to a one-to-one binding model. This calculates a dissociation constant of 150 ± 30 μM.

4.4.2. Competition binding assay

In order to assess the ability of the mimics to competitively inhibit the binding of Gab2b to Grb2 SH3C, a competition assay was conducted with the putative 3_{10} mimics. Each compound was screened at 2.0 mM concentration in duplicate, and the levels of Grb2 SH3C protein detected were scaled relative to samples without compound. Results (Figure 4-6) indicate that none of the mimics are capable of significantly reducing the binding of the Gab2b peptide. The positive control (100x of free peptide) did inhibit the activity, as expected.

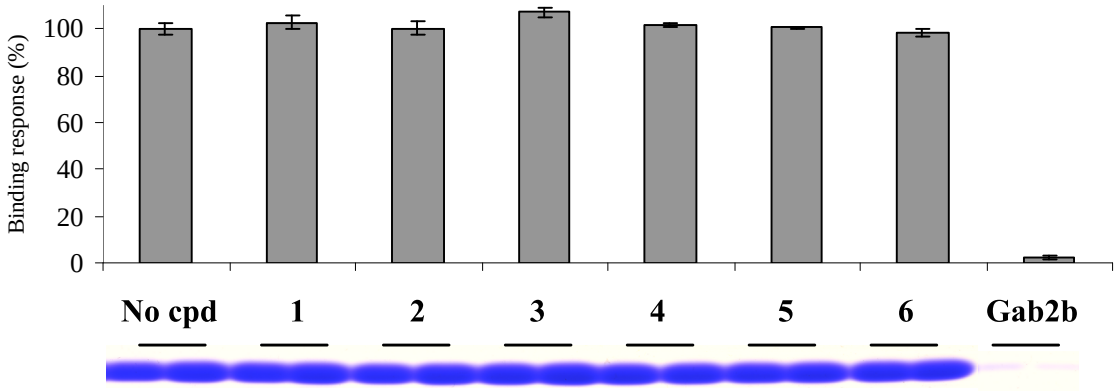


Figure 4-6: Normalized intensities (top) and Coomassie blue-dyed bands (bottom) for the competition assay using the 3_{10} mimics. Mimics were screened at a concentration of 2.0 mM. None showed significant inhibitory effect. The positive control Gab2b peptide is in 100x molar excess.

4.4.3. NMR: ^{15}N HSQC shifts

With SPR and the competition assay providing conflicting information about the binding of the small molecules, the NMR ^{15}N HSQC shift experiment was conducted with mimic **3**. Control experiments (described in Chapter 3) established that the Gab2b peptide causes large NMR shifts in the protein. The residues perturbed in the NMR experiment map very closely to the residues along the crystallographic interface, confirming that the experiment is able to reproduce the solid-state binding of the peptide. In this case, ^{15}N Grb2 SH3C (120 μM, 25 mM phosphate pH 7.2, 15 mM NaCl) was incubated with 20 molar equivalents of mimic **3** (2.4 mM). Compared to the spectrum acquired without ligand, several large shifts are evident (Figure 4-7). There is no selective purification of the protein as seen for some of the PPII mimics, but the magnitude of the HSQC shifts is much greater. Mapping the largest shifts to the crystal structure of Grb2 SH3C bound to the Gab2b peptide (PDB 2VWF)[16] demonstrates that most of the shifted residues are in the core of the protein. The two largest shifts (Ile25 and Phe47) are completely solvent-excluded, with their sidechains pointing into the conserved β-sandwich of the protein. Relatively few shifts occur to the surface residues of the protein, implying that most of the binding is nonspecific unfolding of the domain.

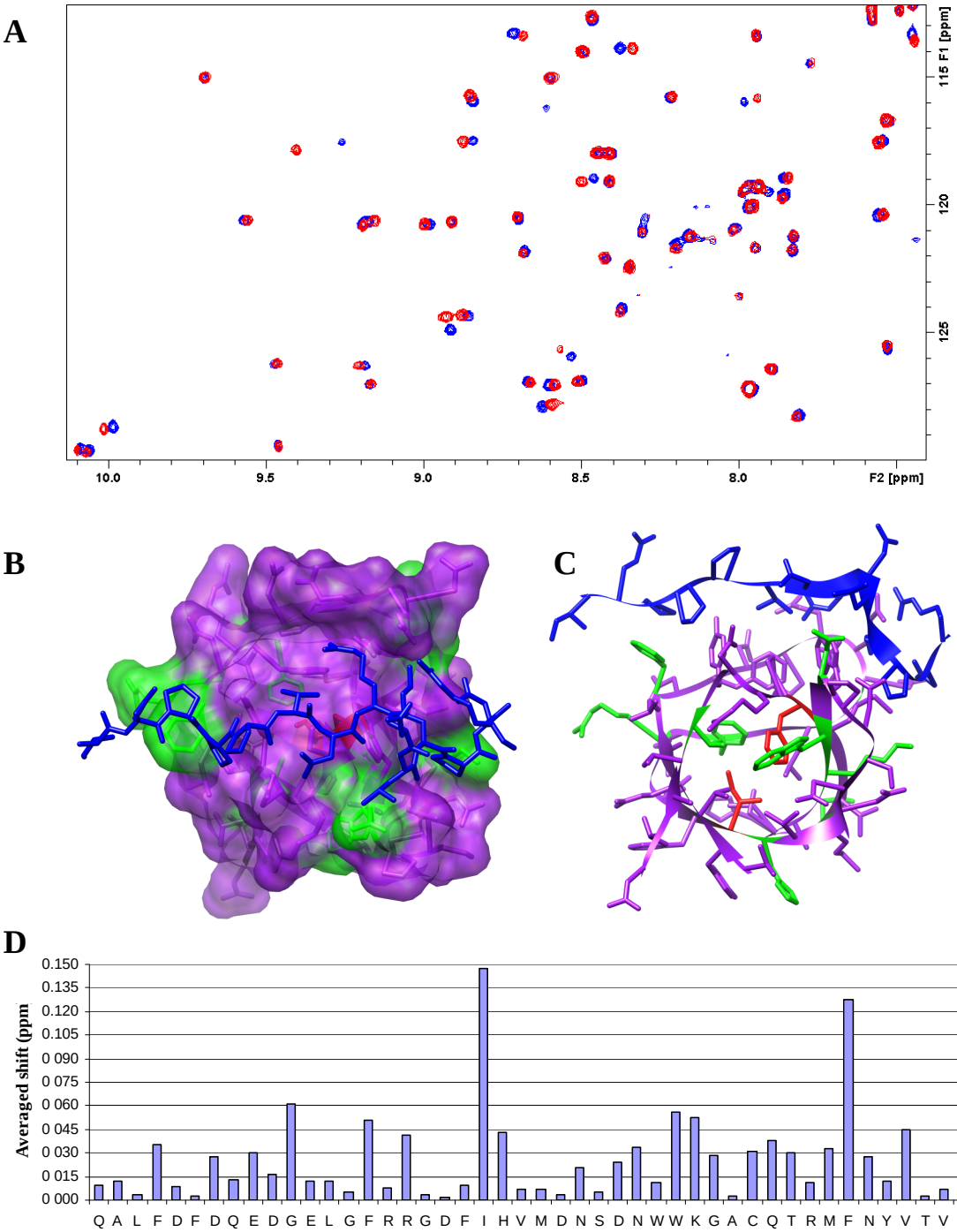


Figure 4-7: (A): ^{15}N HSQC overlay of the apo protein (340 μM, pH 7.2; red) and protein (120 μM, pH 7.2; blue) with 20 eq **3**. Several large chemical shift changes are apparent, although many residues shift very little. (B): model of the interacting residues found by the NMR shift experiments. Residues which shifted > 0.03 ppm are coloured green and mapped onto the crystal structure (2VWF) of Grb2 SH3C binding the Gab2b 3_{10} peptide (blue). (C) view into the core of Grb2 SH3C. The most strongly shifting residues (Ile25 and Phe47) are coloured red. These point into the core of the protein, suggesting that the key change upon binding is an unfolding of the domain. (D): histogram of HSQC shifts for each residue in Grb2 SH3C.

4.5. Discussion

The work presented in this chapter attempted to mimic the 3_{10} helix of the tightly-binding Gab2b peptide. This helix contains the key Arg and Lys residues required for peptide binding, and so it was hoped that a small mimic incorporating the appropriate functionality would be able to bind to Grb2 SH3C. This proved not to be the case. Although SPR suggests that one of the mimics may bind to the domain, the overall response is low and not consistent among two protein constructs. The competition assay likewise failed to show significant binding for any of the molecules. Interestingly, NMR suggests that binding is in fact occurring, but rather than it being a specific interaction it results in a general unfolding of the protein. Not only is this effect undesirable in a mimic, it is unlikely to be specific for Grb2 SH3C. Denaturing of the protein would be expected to decrease binding in the competition assay (by unfolding protein rather than interrupting specific binding), and given the pre-incubation of mimic and protein in that assay it is strange that no response was seen. It is possible that upon addition of the bead-bound peptide, the protein refolds. Alternatively, the response seen in NMR could be the weighted average of mostly folded protein with a small fraction that is reversibly unfolded. In this case, most of the protein would be available to bind the peptide, and perhaps once bound the protein becomes resistant to unfolding by the mimic.

The benzamide scaffold is somewhat too long to perfectly match the 3_{10} helix, and it is possible that the scaffold is thus inherently limited as a mimic for this structure. There has been no successful use of peptidomimetics for inhibiting 3_{10} helix-mediated interactions, so the molecules designed here cannot be compared to others. Alternatively, the mimic may be too small to gain sufficient favourable binding interactions. The incorporation of additional sidechains may therefore cause binding. In the case of Gab2b, such sidechains would have to match the structure of the rest of the peptide, which does not fold in a neat 3_{10} helix. The

minimum peptide length required for binding could potentially be investigated with truncates of the Gab2b sequence, but the point where loss of activity is found could be due to misfolding of a shorter peptide rather than unfavourable binding energies.

Moving forward, significantly larger screening efforts are needed to assess both the utility of the benzamide scaffold and the possibility of such mimics targeting Grb2 SH3C. Large libraries of benzamide-type compounds have been synthesized,[22,24] though these do not include Arg sidechains. The chemistry described here of synthesizing Arg from Lys sidechains creates a molecule that is amenable to additional coupling reactions, as evidenced by the successful synthesis of Arg/Lys/Pro mimics. As a result, these monomers should be incorporated into future screening libraries, which would extend the versatility of these collections. By testing such a library against Grb2 SH3C, it could be determined whether any benzamide is capable of binding to the domain. Alternatively, the same collections (or even the subset of molecules synthesized here) could be tested against multiple SH3 domains as a means of probing the selectivity of these proteins.

4.6. Conclusion

The synthesis of benzamide mimics containing cationic sidechains was completed. These molecules incorporate Arg sidechains into this scaffold for the first time, and are synthesized using orthogonal azide and Boc protecting groups. A small library of such compounds was synthesized, and found not to bind significantly to Grb2 SH3C. NMR evidence suggests that one of the molecules causes at least partial unfolding of the domain, although this does not appear to stop the binding of the Gab2b peptide. Screening of these mimics against other SH3 domains may help elucidate whether this response is specific to Grb2 SH3C or if cationic benzamides are generally deleterious to protein structure.

4.7. Perspective on small molecule binders to Grb2 SH3C

The field of peptidomimetics assumes that complex protein-protein interactions can be inhibited by a small scaffold displaying relatively few sidechains. This approach has proved remarkably successful against a range of targets, in some cases with multiple scaffolds able to bind to the same protein with similar affinity provided the correct sidechains are presented. This has been the case with Bcl-x_L, which has been successfully bound with the terphenyl,[15,27] oligopyridylamide,[28] terephthalamide,[29] benzoylurea,[30] biphenyl,[31] and "wet edge"[32] scaffolds (Figure 4-8). These mimics incorporate hydrophobic sidechains, and bind to a relatively solvent-excluded pocket on Bcl-x_L which natively binds a rigid, hydrophobic helix. Grb2 SH3C, however, provides three challenges for peptidomimetic design: first, it is a signalling molecule, with weak affinities even for the native ligands. This may limit the maximum affinities that can be achieved by small molecules since the protein has evolved to avoid very tight binding of any ligand. Second, the native ligands have very different PPII and 3_{10} structures, neither of which has been mimicked with a small synthetic molecule. Finally, the key binding residues are hydrophilic, which introduces potential desolvation penalties in binding a ligand.[33]

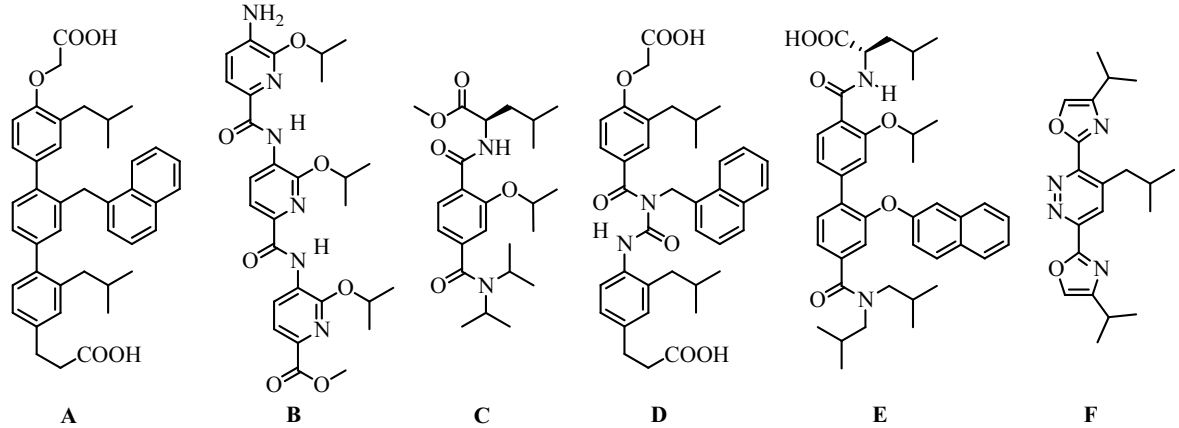


Figure 4-8: Different peptidomimetics which bind to Bcl-x_L. (a) terphenyl, K_i 0.114 μM. (b) oligopyridylamide, K_i 2.3 μM. (c) terephthalamide, K_i 0.781 μM. (d) benzoylurea, K_i 2.4 μM. (e) biphenyl, K_i 1.8 μM. (f) "wet-edge" mimic, K_i not determined.

Two classes of peptidomimetics were synthesized in a search for binders to Grb2 SH3C. Neither class was overwhelmingly successful, and while SPR and NMR indicate that several of the molecules bind to the domain, this binding is relatively nonspecific and was not successfully quantified using the biophysical or biochemical methods tested. This raises general concerns regarding rational design approaches. When successful, rationally-designed molecules can target the key structural features of a particular target. This avoids the cost and difficulties associated with large-scale screening, and may produce molecules that are highly selective for their targets. On the other hand, when unsuccessful or when steep SAR is found, the designed molecules do not necessarily provide much explanation as to why they failed. It could be that the design is close, and that a minute change to the scaffold may result in an active molecule. Alternatively, the molecules may be intrinsically unsuitable, and novel scaffolds or approaches may be required. Differentiating between these options is challenging, and thus the only route forward may be to synthesize and test additional molecules. *In extremis*, this results in the screening of large combinatorial libraries, as has indeed taken place within the field.[22,24]

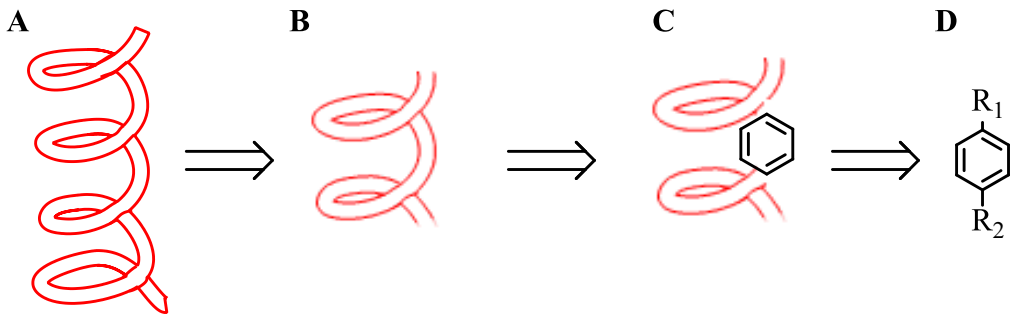


Figure 4-9: Potential theoretical progression from (a) peptide ligand to (b) peptide truncate to (c) peptide/peptidomimetic hybrid to (d) fully-synthetic peptidomimetic.

One strategy for bridging the large gap between peptide ligands and peptidomimetics may be to systematically probe the ligand chemical space (Figure 4-9). For instance, in an initial approach the sequence of the native peptide could be varied, for instance in a spot blot assay. This may reveal which residues are critical for binding, as well as the tolerance

of the site for mutations (e.g. in the size of a hydrophobic residue). This could lead to peptide truncates and the screening of small libraries of mutant peptides, which could define a minimal number of interacting residues. Very small peptides may not fold, however, but these could be stabilized by approaches such as stapling.[34] With a better idea of the binding site, hybrid mimic/peptide molecules could be created which eliminate selected residues of the native sequence. This approach has been used to discover inhibitors of the Fyn SH3 domain using a Pro-Pro mimic within a native PPII polypeptide.[35] Particularly for scaffolds such as the benzamide, this chemistry could conceivably be conducted using entirely solid-phase methods. Finally, a smaller number of wholly-synthetic peptidomimetics could be synthesized, taking advantage of the structural information gleaned from these studies.

The lack of published inhibitors of Grb2 SH3C complicates ligand design enormously, and increases the potential rewards for a successful molecule. Principally, it has not yet been established whether the interaction is "druggable" with a small molecule,[36] and thus whether inactive compounds result from poor design or a generally intractable target. Unfortunately, the results found with the screened molecules in Chapter 2 and the designed molecules in Chapters 3 & 4 do not conclusively answer this question. Encouragingly, several molecules have been identified which clearly bind to the Gab2-binding groove of Grb2 SH3C, but all do so in a multipoint manner. Selectivity of these molecules for Grb2 SH3C, as opposed to other SH3 domains or even very different proteins, remains to be explored.

Figure removed for
copyright reasons

Figure 4-10: Flexibility of Grb2 SH3C to bind the structurally-diverse Gab2a and Gab2b peptides. Crystal structures of the complex (left) show that the peptides contact similar residues on the protein, despite very different orientations of the key Arg and Lys sidechains (right). Adapted from Harkiolaki *et al.*[16]

Given the flexibility of Grb2 SH3C to bind such structurally-diverse peptides as Gab2a and Gab2b (Figure 4-10), it is possible that the domain is simply too flexible to accommodate small molecules in a single, well-defined binding site. If this is the case, a molecule might have to be relatively large to bind to a single site on the protein. The benzamide molecules, which lack significant hydrophobic bulk, are certainly less potent binders then the benzoylureas, but this might be due to the different projection of sidechains in the scaffold rather than hydrophobic properties. Likewise, greater activity was found for the dihydro-*s*-triazine screened molecules which incorporated larger hydrophobic bulk, but even this was not sufficient to allow the molecules to bind stoichiometrically. Moreover, the trimeric benzoylurea bound with less specificity than the smaller mimics, despite the increased hydrophobicity of that molecule. What is clear, however, is a strong preference for Grb2 SH3C binding cationic molecules, with no neutral molecules found to bind with any noticeable activity. Incorporating additional hydrogen bonding capability, for instance,

as seen on the Arg/Arg molecules, is not a panacea, as these molecules did not bind significantly more strongly by SPR and may be less selective than the mixed Arg/Lys mimics. Spot blot analysis with the native Gab2b peptide indicates that mutation of Lys to Arg is not tolerated, but this should be confirmed by *in vitro* testing of the mutated peptide.

Moving forward, additional screening is required to further probe binding to Grb2 SH3C. Computationally, screens might include larger molecules outside the "Lipinski space" which may be better for inhibiting PPIs. Additionally, virtual screens in which no hydrogen bond constraint is applied might help discover novel binding modes to the protein. Now that tractable SPR and NMR conditions for investigating binding to Grb2 SH3C have been determined, either of these methods could be used in fragment-based *in vitro* screens for small molecule binders. Future *in vitro* screens should also be expanded to include additional RxxK SH3 domains. By cross-screening molecules against multiple protein targets, some idea of the selectivity of each binding site may become apparent, which could lead to more-selective therapies. The large number of SH3 domains in the genome make finding absolute selectivity unlikely, but if a peptidomimietic binds to only a subset of these proteins, it could allow the development of a pharmacophore model for their binding sites. This would be useful in the synthesis of additional peptidomimetics, and in guiding virtual screening.

4.8. Experimental details

4.8.1. SPR

This assay was performed according to the general SPR methods described in Chapter 3. Measurements were taken according to the procedure for 2nd generation mimics described in Chapter 3.

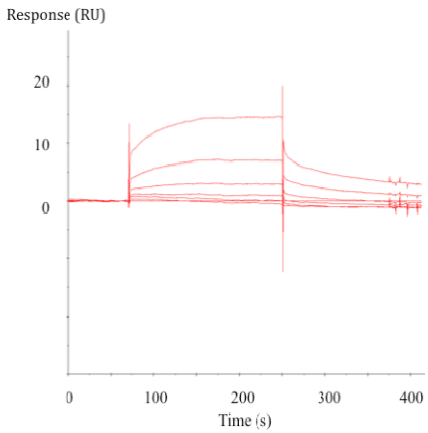


Figure 4-11: Sensograms of mimic **3** binding to ^{15}N-labeled Grb2 SH3C. Mimic concentrations are a two-fold dilution series from 200 μM (highest response) to 6.25 μM (lowest non-zero response). The lowest mimic concentrations yield sensograms that are to similar to be analyzed.

4.8.2. Competition binding assay

This assay was performed according to the procedure described in Chapter 3.

4.8.3. NMR

This assay was performed according to the procedure described in Chapter 3, with the exception that spectra of ligand-bound ^{15}N Grb2 SH3C were obtained at a protein concentration of 120 μM in a buffer of 25 mM phosphate pH 7.2 + 15 mM NaCl + 2 mM DTT in 90:20 H_2O:D_2O.

4.8.4. Synthesis

General synthetic details and methods are described in Chapter 3. The spectroscopic details of all compounds synthesized in this chapter are given below.

4.9. References

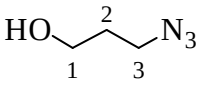
(1) Donohue, J. Hydrogen Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A.* **1953**, *39*, 470–478.
(2) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205–211.
(3) Barlow, D. J.; Thornton, J. M. Helix geometry in proteins. *J. Mol. Biol.* **1988**, *201*, 601–619.
(4) Pal, L.; Basu, G. Novel protein structural motifs containing two-turn and longer 3(10)-helices. *Protein Eng.* **1999**, *12*, 811–814.

(5) Enkhbayar, P.; Hikichi, K.; Osaki, M.; Kretsinger, R. H.; Matsushima, N. 310-helices in proteins are parahelices. *Proteins Struct. Funct. Bioinforma.* **2006**, *64*, 691–699.
(6) Karle, I. L.; Balaram, P. Structural characteristics of .alpha.-helical peptide molecules containing Aib residues. *Biochemistry (Mosc.)* **1990**, *29*, 6747–6756.
(7) Basu, G.; Bagchi, K.; Kuki, A. Conformational preferences of oligopeptides rich in α-aminoisobutyric acid. I. Observation of a 310/α-helical transition upon sequence permutation. *Biopolymers* **1991**, *31*, 1763–1774.
(8) Schievano, E.; Bisello, A.; Chorev, M.; Bisol, A.; Mammi, S.; Peggion, E. Aib-Rich Peptides Containing Lactam-Bridged Side Chains as Models of the 310-Helix. *J. Am. Chem. Soc.* **2001**, *123*, 2743–2751.
(9) Blackwell, H. E.; Grubbs, R. H. Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis. *Angew. Chem. Int. Ed.* **1998**, *37*, 3281–3284.
(10) Boal, A. K.; Guryanov, I.; Moretto, A.; Crisma, M.; Lanni, E. L.; Toniolo, C.; Grubbs, R. H.; O'Leary, D. J. Facile and E-Selective Intramolecular Ring-Closing Metathesis Reactions in 310-Helical Peptides: A 3D Structural Study. *J. Am. Chem. Soc.* **2007**, *129*, 6986–6987.
(11) Jacobsen, Ø.; Maekawa, H.; Ge, N.-H.; Görbitz, C. H.; Rongved, P.; Ottersen, O. P.; Amiry-Moghaddam, M.; Klaveness, J. Stapling of a 310-Helix with Click Chemistry. *J. Org. Chem.* **2011**, *76*, 1228–1238.
(12) Ko, E.; Liu, J.; Perez, L. M.; Lu, G.; Schaefer, A.; Burgess, K. Universal Peptidomimetics. *J. Am. Chem. Soc.* **2011**, *133*, 462–477.
(13) Raghuraman, A.; Ko, E.; Perez, L. M.; Ioerger, T. R.; Burgess, K. Pyrrolinone–Pyrrolidine Oligomers as Universal Peptidomimetics. *J. Am. Chem. Soc.* **2011**, *133*, 12350–12353.
(14) Mammen, M.; Shakhnovich, E. I.; Whitesides, G. M. Using a Convenient, Quantitative Model for Torsional Entropy To Establish Qualitative Trends for Molecular Processes That Restrict Conformational Freedom. *J. Org. Chem.* **1998**, *63*, 3168–3175.
(15) Yin, H.; Lee, G.-I.; Sedey, K. A.; Kutzki, O.; Park, H. S.; Orner, B. P.; Ernst, J. T.; Wang, H.-G.; Sebti, S. M.; Hamilton, A. D. Terphenyl-Based Bak BH3 alpha-helical proteomimetics as low-molecular-weight antagonists of Bcl-xL. *J. Am. Chem. Soc.* **2005**, *127*, 10191–10196.
(16) Harkiolaki, M.; Tsirka, T.; Lewitzky, M.; Simister, P. C.; Joshi, D.; Bird, L. E.; Jones, E. Y.; O'Reilly, N.; Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the SH3C Domain of Grb2. *Structure* **2009**, *17*, 809–822.
(17) Saraogi, I.; Incarvito, C. D.; Hamilton, A. D. Controlling Curvature in a Family of Oligoamide α-Helix Mimetics. *Angew. Chem.* **2008**, *120*, 9837–9840.
(18) Zhang, D.-W.; Zhao, X.; Hou, J.-L.; Li, Z.-T. Aromatic Amide Foldamers: Structures, Properties, and Functions. *Chem. Rev.* **2012**, *112*, 5271–5316.
(19) Plante, J.; Campbell, F.; Malkova, B.; Kilner, C.; Warriner, S. L.; Wilson, A. J. Synthesis of functionalised aromatic oligamide rods. *Org. Biomol. Chem.* **2008**, *6*, 138.
(20) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601–2611.
(21) Ahn, J.-M.; Han, S.-Y. Facile synthesis of benzamides to mimic an α-helix. *Tetrahedron Lett.* **2007**, *48*, 3543–3547.
(22) Shaginian, A.; Whitby, L. R.; Hong, S.; Hwang, I.; Farooqi, B.; Searcey, M.; Chen, J.; Vogt, P. K.; Boger, D. L. Design, Synthesis, and Evaluation of an α-Helix

Mimetic Library Targeting Protein−Protein Interactions. *J. Am. Chem. Soc.* **2009**, *131*, 5564–5572.
(23) Kulikov, O. V.; Hamilton, A. D. Synthesis of the novel trimeric benzamides—potential inhibitors of protein–protein interactions. *RSC Adv.* **2012**, *2*, 2454.
(24) Murphy, N. S.; Prabhakaran, P.; Azzarito, V.; Plante, J. P.; Hardie, M. J.; Kilner, C. A.; Warriner, S. L.; Wilson, A. J. Solid-Phase Methodology for Synthesis of O-Alkylated Aromatic Oligoamide Inhibitors of α-Helix-Mediated Protein–Protein Interactions. *Chem. – Eur. J.* **2013**, *19*, 5546–5550.
(25) Stewart, J. J. P. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *J. Mol. Model.* **2004**, *10*, 155–164.
(26) Kulikov, O. V.; Thompson, S.; Xu, H.; Incarvito, C. D.; Scott, R. T. W.; Saraogi, I.; Nevola, L.; Hamilton, A. D. Design and Synthesis of Oligoamide-Based Double α-Helix Mimetics. *Eur. J. Org. Chem.* **2013**, *2013*, 3433–3445.
(27) Kutzki, O.; Park, H. S.; Ernst, J. T.; Orner, B. P.; Yin, H.; Hamilton, A. D. Development of a Potent Bcl-xL Antagonist Based on α-Helix Mimicry. *J. Am. Chem. Soc.* **2002**, *124*, 11838–11839.
(28) Ernst, J. T.; Becerril, J.; Park, H. S.; Yin, H.; Hamilton, A. D. Design and Application of an α-Helix-Mimetic Scaffold Based on an Oligoamide-Foldamer Strategy: Antagonism of the Bak BH3/Bcl-xL Complex. *Angew. Chem. Int. Ed.* **2003**, *42*, 535–539.
(29) Yin, H.; Hamilton, A. D. Terephthalamide derivatives as mimetics of the helical region of Bak peptide target Bcl-xL protein. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1375–1379.
(30) Rodriguez, J. M.; Hamilton, A. D. Benzoylurea Oligomers: Synthetic Foldamers That Mimic Extended α Helices. *Angew. Chem. Int. Ed.* **2007**, *46*, 8614–8617.
(31) Rodriguez, J. M.; Nevola, L.; Ross, N. T.; Lee, G.; Hamilton, A. D. Synthetic Inhibitors of Extended Helix–Protein Interactions Based on a Biphenyl 4,4′-Dicarboxamide Scaffold. *ChemBioChem* **2009**, *10*, 829–833.
(32) Biros, S. M.; Moisan, L.; Mann, E.; Carella, A.; Zhai, D.; Reed, J. C.; Rebek Jr., J. Heterocyclic α-helix mimetics for targeting protein–protein interactions. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4641–4645.
(33) Biela, A.; Khayat, M.; Tan, H.; Kong, J.; Heine, A.; Hangauer, D.; Klebe, G. Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. *J. Mol. Biol.* **2012**, *418*, 350–366.
(34) Kim, Y.-W.; Grossmann, T. N.; Verdine, G. L. Synthesis of all-hydrocarbon stapled α-helical peptides by ring-closing olefin metathesis. *Nat. Protoc.* **2011**, *6*, 761–771.
(35) Li, H.; Lawrence, D. S. Acquisition of Fyn-Selective SH3 Domain Ligands via a Combinatorial Library Strategy. *Chem. Biol.* **2005**, *12*, 905–912.
(36) Owens, J. Determining druggability. *Nat. Rev. Drug Discov.* **2007**, *6*, 187–187.
(37) Badiang, J. G.; Aubé, J. One-Step Conversion of Aldehydes to Oxazolines and 5,6-Dihydro-4H-1,3-oxazines Using 1,2- and 1,3-Azido Alcohols. *J. Org. Chem.* **1996**, *61*, 2484–2487.

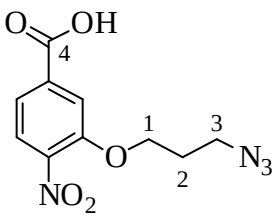
4.10. Experimental Data for Synthesized Compounds

3-Azidopropan-1-ol (7)



Sodium azide (1.95 g, 30 mmol, 1.05 eq) was dissolved in DMSO (60 mL). To this solution was added 3-bromo-1-propanol (3.97 g, 28.6 mmol, 1.0 eq) and the mixture heated to 70 ºC and stirred overnight. The reaction was partitioned between water (100 mL) and ether (100 mL) and washed with ether (3 x 60 mL). The organic layers were combined, washed with brine, dried ($MgSO_4$), filtered, and concentrated at 500 mBar and 40 ºC. Purification *via* flash column chromatography (1:1 ether:petrol) yielded the title compound as a volatile, colourless oil (2.70 g, 26.7 mmol, 93 %): δ_H (500 MHz, $CDCl_3$) 3.82 (2H, q, *J* 5.7, H1), 3.53 (2H, quint, *J* 3.5, H3), 1.90 (2H, quint, *J* 6.3, H2). δ_C (100 MHz, $CDCl_3$) 60.0 (C1), 48.5 (C3), 31.4 (C2). IR υ_{max} 3335, 2946, 2882, 2090, 1045. HRMS (FI) found 101.0590. $C_3H_7N_3O^+$ $[M]^+$ requires 101.0590. Spectroscopic measurements agreed with literature values.[37]

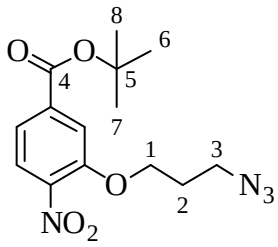
3-(3-Azidopropoxy)-4-nitrobenzoic acid (9)



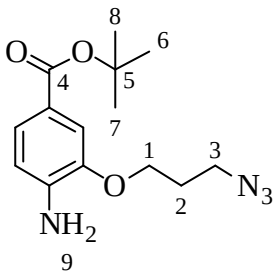
Following the procedure of Boger *et al*,[22] azide **7** (655 mg, 6.5 mmol, 1.2 eq) was added dropwise to a stirred solution of sodium hydride (60 % dispersion in oil, 540 mg, 13.5 mmol, 2.5 eq) in THF (30 mL) at 0 ºC. The mixture was stirred for 15 minutes at 0 ºC before 3-fluoro-4-nitrobenzoic acid **8** (1.00 g, 5.4 mmol, 1.0 eq) was added. The reaction was stirred for 5 minutes at 0 ºC, then 2 hours at room temperature, before being diluted with EtOAc (20 mL). NH_4Cl (20 mL) was added, and the mixture was washed with HCl (0.5 M, 20 mL x 3). The organic layer was collected, dried ($MgSO_4$), filtered, and concentrated, and the product purified *via* flash column chromatography (3:2 EtOAc:Pet + 1 % AcOH) to yield the title compound as a yellow solid (1.40 g, 5.26 mmol, 97 %): MP

123-126. δ_H (400 MHz, MeOD) 7.86 (1H, d, *J* 10.9, Ar), 7.85 (1H, s, Ar), 7.72 (1H, dd, *J* 8.3; 1.6, Ar), 4.30 (2H, t, *J* 5.9, H1), 3.56 (2H, t, *J* 6.6, H3), 2.09 (2H, quint, *J* 6.2, H2). δ_C (125 MHz, MeOD) 166.6 (C4), 151.6 (Ar), 143.0 (Ar), 136.0 (Ar), 125.1 (Ar), 121.8 (Ar), 115.6 (Ar), 66.6 (C1), 28.5 (C2). IR υ_{max} 3061, 2924, 2853, 2102, 1690, 1589, 1530, 1259. HRMS (ESI) found 265.0569 $C_{10}H_9N_4O_5^-$ $[M-H]^-$ requires 265.0578.

***tert*-Butyl 3-(3-azidopropoxy)-4-nitrobenzoate (10)**



Carboxylic acid **9** (1.4 g, 5.26 mmol, 1.0 eq) was dissolved in DCM (30 mL) and cooled to 0 ºC and DMAP (96 mg, 0.79 mmol, 0.15 eq), *tert*-butanol (2 mL, 21 mmol, 4.0 eq), and *N*,*N'*-dicyclohexylcarbodiimide (1.25 g, 6 mmol, 1.15 eq) were added. The mixture was warmed to room temperature and stirred for 68 hours before 10 % NaOH (100 mL) and DCM (100 mL) were added. The organic layer was washed with water (2 x 100 mL) and brine (1 x 100 mL), then dried ($MgSO_4$) and purified *via* flash column chromatography (10:1 petrol:ether) to yield the title compound as a yellow solid (952 mg, 2.95 mmol, 56 %): MP 74-75. δ_H (400 MHz, $CDCl_3$) 7.78 (1H, d, *J* 8.36, Ar), 7.67 (1H, d, *J* 1.3, Ar), 7.60 (1H, dd, *J* 8.3; 1.5, Ar), 4.23 (2H, t, *J* 5.8, H1), 3.65 (1H, t, *J* 6.5, H3), 2.07 (2H, quint, *J* 6.1, H2), 1.58 (9H, s, H 6-8). δ_C (100 MHz, $CDCl_3$) 163.6 (C4), 151.5 (Ar), 142.0 (Ar), 136.9 (Ar), 125.2 (Ar), 121.4 (Ar), 115.3 (Ar), 82.5 (C5), 66.1 (C1), 47.7 (C3), 28.4 (C2), 28.0 (C6-8). IR υ_{max} 3113, 3079, 2986, 2086, 1719, 1611, 1521, 1249. HRMS (ESI) found 345.1163 $C_{14}H_{18}N_4O_5Na^+$ $[M+Na]^+$ requires 345.1169.



***tert*-Butyl 4-amino-3-(3-azidopropoxy)benzoate (11)**

Iron powder (1.65 g, 29.5 mmol, 10 eq) and ammonium chloride (79 mg, 1.48 mmol, 0.5 eq) were added to a stirred solution of the

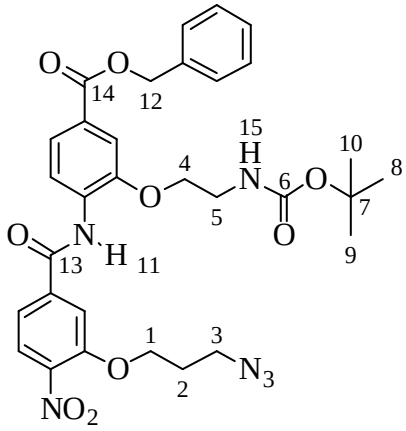
aromatic nitro compound **10** (952 mg, 2.95 mmol) dissolved in ethanol/water (40 mL:10 mL). The mixture was heated to 80 ºC for 30 minutes then filtered through Celite® with EtOAc (150 mL). Solvent was removed *in vacuo* and the product purified *via* flash column chromatography (3:2 petrol:ether) to yield the title compound as a colorless oil (560 mg, 1.92 mmol, 65 %): δ_H (250 MHz, $CDCl_3$) 7.48 (1H, dd, *J* 8.1; 1.7, Ar), 7.40 (1H, d, *J* 1.7, Ar), 6.62 (1H, d, *J* 8.2, Ar), 4.24 (2H, br s, H9), 4.08 (2H, t, *J* 6.0, H1), 3.46 (2H, t, *J* 6.6, H3), 2.04 (2H, quint, *J* 6.3, H2), 1.55 (9H, s, H6-8). δ_C (62.5 MHz, $CDCl_3$) 166.6 (C4), 145.4 (Ar), 141.4 (Ar), 124.5 (Ar), 121.7 (Ar), 113.7 (Ar), 112.6 (Ar), 80.6 (C5), 65.3 (C1), 48.8 (C3), 29.0 (C2), 28.7 (C6-8). IR υ_{max} 3488, 3369, 2976, 2932, 2098, 1691, 1617, 1297. HRMS (ESI) found 315.1425 $C_{14}H_{20}N_4O_3Na^+$ $[M+Na]^+$ requires 315.1428.

4.10.1. General Procedure for Amidation

DIPEA (2.5 eq) was added dropwise to a solution of carboxylic acid (1.0 eq) in DCM. Mukaiyama's reagent (2-chloro-1-methylpyridinium iodide, 1.2 eq) was added to the mixture and the reaction heated to 40 ºC for 30 minutes. Amine (1.1 eq) was then added and the reaction stirred for 20 hours at 40 ºC. The solution was diluted with NH_4Cl (10 mL.100 mg^{-1}) and DCM (10 mL.100 mg^{-1}) and extracted with DCM (3 x 10 mL.100 mg^{-1}). The organic layers were combined, dried (Na_2SO_4), filtered, and purified *via* flash column chromatography.

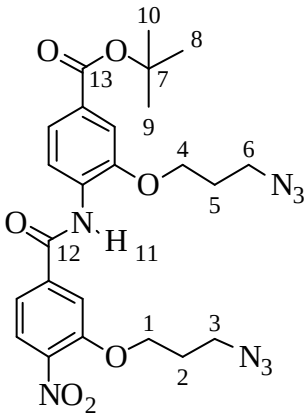
Benzyl-4-(3-(3-azidopropoxy)-4-nitrobenzamido)-3-(2-((*tert*-butoxycarbonyl)amino)ethoxy) benzoate (14)

Following the general procedure for amidation, acid **9** (71 mg, 0.27 mmol) was coupled with benzyl 4-amino-3-(2-((*tert*-butoxycarbonyl)amino)ethoxy)benzoate (114 mg, 0.30 mmol) and purified (1:1 petrol:ether) to yield the title compound as a yellow solid (161 mg, 0.25 mmol, 95 %): MP 105-108. δ_H (400 MHz, $CDCl_3$) 9.19 (1H, s, H11), 8.64 (1H, d, *J*



9.5, Ar), 7.89 (1H, d, *J* 8.4, Ar), 7.80-7.74 (2H, m, Ar), 7.69 (1H, d, *J* 8.2, Ar), 7.52 (1H, s, Ar), 7.47-7.33 (5H, m, Ar), 5.35 (2H, s, H12), 4.97 (1H, t, *J* 6.6, H15), 4.23 (2H, t, *J* 5.7, H1), 4.11 (2H, t, *J* 4.9, H4), 3.62 (2H, q, *J* 5.4, H5), 3.57 (2H, t, *J* 6.6, H3), 2.10 (2H, quint, *J* 6.3, H2), 1.18 (9H, s, H8-10). δ_C (100 MHz, $CDCl_3$) 165.9 (C13 or C14), 164.1 (C13 or C14), 156.5 (C6), 151.8 (Ar), 147.2 (Ar), 141.5 (Ar), 139.6 (Ar), 136.0 (Ar), 132.0 (Ar), 128.6 (Ar), 128.5 (Ar), 128.3 (Ar), 128.0 (Ar), 125.5 (Ar), 123.6 (Ar), 119.6 (Ar), 119.1 (Ar), 114.6 (Ar), 111.1 (Ar), 79.6 (C7), 69.6 (C4), 66.8 (C12), 66.2 (C1), 47.8 (C3), 40.6 (C5), 28.4 (C2), 28.0 (C8-10). IR υ_{max} 3369, 2934, 2252, 1704, 1601, 1522, 1269. HRMS (ESI) found 657.2284 $C_{31}H_{34}N_6O_9Na^+$ $[M+Na]^+$ requires 657.2279.

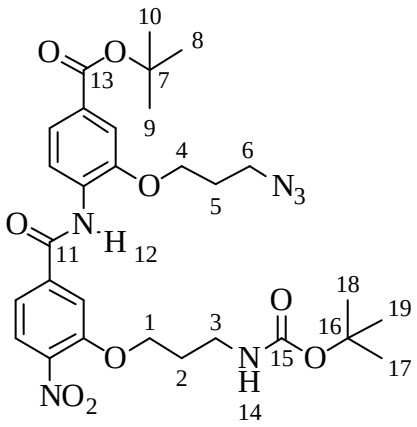
***tert*-Butyl 3-(3-azidopropoxy)-4-(3-(3-azidopropoxy)-4-nitrobenzamido)benzoate (15)**



Following the general procedure for amide formation, acid **9** (100 mg, 0.38 mmol, 1.0 eq) was coupled with amine **11** (115 mg, 0.39 mmol, 1.05 eq) and the product purified (3:2 ether:petrol) to yield the title compound (193 mg, 0.36 mmol, 95 %) as a yellow solid: MP 128-129. δ_H (250 MHz, $CDCl_3$) 8.78 (1H, s, H11), 8.55 (1H, d, *J* 8.5, Ar), 7.94 (1H, d, *J* 8.3, Ar), 7.73-7.67 (2H, m, Ar), 7.56 (1H, d, *J* 1.6, Ar), 7.44 (1H, dd, *J* 8.4, 1.6, Ar), 4.34-4.23 (4H, m, H1 and H4), 3.63-3.50 (4H, m, H3 and H6), 2.22-2.06 (4H, m, H2 and H5), 1.60 (9H, s, H8-10). δ_C (62.5 MHz, $CDCl_3$) 165.6 (C12 or C13), 163.6 (C12 or C13), 152.7 (Ar), 147.2 (Ar), 141.9 (Ar), 140.4 (Ar), 131.5 (Ar), 128.3 (Ar), 126.4 (Ar), 123.8 (Ar), 119.4 (Ar), 118.2 (Ar), 114.6 (Ar), 112.0 (Ar), 81.7 (C7), 67.1 (C1 or C4), 66.8 (C1 or C4), 49.4 (C3 or C6), 48.1 (C3 or C6), 28.9 (C2 or C5), 28.8 (C2 or C5), 28.6 (C8-10). IR υ_{max} 3427, 2933, 2094,

1706, 1601, 1521, 1164. HRMS (ESI) found 539.2002 $C_{24}H_{27}N_8O_7^+$ $[M+H]^+$ requires 539.2008.

tert-Butyl 3-(3-azidopropoxy)-4-(3-(3-((_tert_-butoxycarbonyl)amino)propoxy)-4-nitrobenzamido) benzoate (16)



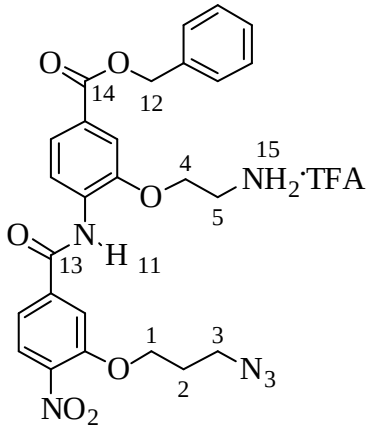
Following the general procedure for amidation, 3-(3-((*tert*-butoxycarbonyl)amino)propoxy)-4-nitrobenzoic acid (443 mg, 1.30 mmol, 1.0 eq) was coupled to amine **11** (400 mg, 1.37 mmol, 1.05 eq) and purified (2:1 ether:petrol) to yield the title compound (651 mg, 1.06 mmol, 81 %) as a white solid: MP 147-148. δ_H (300 MHz, $CDCl_3$) 8.75 (1H, s, H12), 8.47 (1H, d, *J* 8.5, Ar), 7.87 (1H, d, *J* 8.3, Ar), 7.67-7.60 (2H, m, Ar), 7.51 (1H, s, Ar), 7.38 (1H, d, *J* 8.3, Ar), 5.08 (1H, s, H14), 4.26-4.18 (4H, m, H1 and H4), 3.51 (2H, t, *J* 6.1, H6), 3.31 (2H, q, *J* 5.8, H3), 2.12 (2H, quint, *J* 6.1, H5), 2.02 (2H, quint, *J* 5.8, H2), 1.56 (9H, s, H8-10), 1.38 (9H, s, H17-19). δ_C (75 MHz, $CDCl_3$) 165.2 (C11 or C13), 163.2 (C11 or C13), 156.1 (C15), 152.4 (Ar), 146.8 (Ar), 141.4 (Ar), 140.0 (Ar), 131.0 (Ar), 127.9 (Ar), 126.0 (Ar), 123.3 (Ar), 119.0 (Ar), 117.7 (Ar), 114.2 (Ar), 111.5 (Ar), 81.3 (C7), 79.1 (C16), 68.0 (C1 or C4), 66.6 (C1 or C4), 48.9 (C6), 37.8 (C3), 29.0 (C2), 28.5 (C5), 28.4 (C17-19), 28.2 (C8-10). IR υ_{max} 3434, 3338, 2978, 2101, 1711, 1681, 1519, 1162. HRMS (ESI) found 637.2603 $C_{29}H_{38}N_6O_9Na^+$ $[M+Na]^+$ requires 637.2592.

4.10.2. General Procedure for N-Boc Deprotection

TFA (1.0 mL) was added to a stirred solution of an *N*-Boc protected molecule in DCM (1 mL). The mixture was stirred at room temperature and toluene (5 mL) was added. The

solvent was removed *in vacuo* and the resulting compound azeotroped three times with chloroform.

2-(2-(3-(3-azidopropoxy)-4-nitrobenzamido)-5-((benzyloxy)carbonyl)phenoxy) ethanaminium •TFA (17)

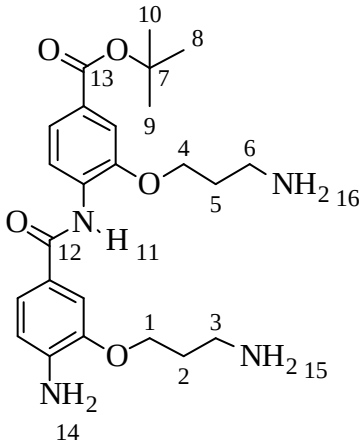


Following the general procedure for *N*-Boc deprotection, protected amine **14** (161 mg, 0.25 mmol) was deprotected to yield the title compound as a yellow solid, which was used for subsequent reactions without further purification (151 mg, 0.24 mmol, 94 %): MP 157-158. δ_H (500 MHz, $CDCl_3$) 9.05 (1H, s, H11), 8.16 (1H, d, *J* 7.8, Ar), 8.07 (3H, s, H15), 7.70 (1H, d, *J* 8.7, Ar), 7.65 (1H, d, *J* 8.3, Ar), 7.53 (1H, s, Ar), 7.49 (1H, s, Ar), 7.40-7.31 (4H, m, Ar), 7.20-7.16 (2H, m, Ar), 5.27 (2H, s, H12), 4.29 (2H, s, H4), 4.14-4.09 (2H, m, H1), 3.45-3.36 (4H, m, H3 and H5), 1.96 (2H, quint, *J* 6.0, H2). δ_C (125 MHz, $CDCl_3$) 166.1 (C14), 164.8 (C13), 161.4 (q, *J* 38, CF_3CO_2), 151.9 (Ar), 147.5 (Ar), 141.1 (Ar), 139.4 (Ar), 135.5 (Ar), 131.0 (Ar), 128.6 (Ar), 128.2 (Ar), 128.0 (Ar), 126.7 (Ar), 125.6 (Ar), 123.9 (Ar), 121.9 (Ar), 118.8 (Ar), 115.6 (q, *J* 289, CF_3CO_2), 114.3 (Ar), 112.0 (Ar), 67.2 (C12), 66.4 (C1), 64.5 (C4), 47.6 (C3), 39.4 (C5), 28.1 (C2). IR υ_{max} 3034, 2955, 2254, 1779, 1525, 1284, 1202. HRMS (ESI) found 535.1946 $C_{26}H_{27}N_6O_7^+$ $[M+H]^+$ requires 535.1936.

4.10.3. General Procedure for Catalytic Hydrogenation

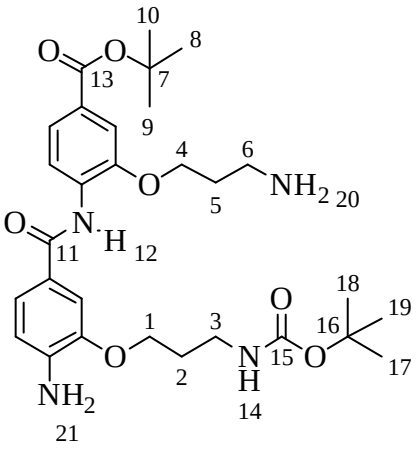
Palladium on activated carbon (10 % by mass; 10 % loading) was added to a stirred solution of an organic molecule in EtOH. The solution was flushed three times with hydrogen gas then allowed to stir under an atmosphere of hydrogen overnight. The solution was filtered through Celite® with DCM (50 $mL.mmol^{-1}$) and the solvent removed *in vacuo*.

***tert*-Butyl 4-(4-amino-3-(3-aminopropoxy)benzamido)-3-(3-aminopropoxy)benzoate (18)**



Following the general procedure for catalytic hydrogenation, benzamide **15** (187 mg, 0.35 mmol) was hydrogenated to yield the title compound (185 mmol, 0.40 mmol, *quant.*) as a viscous yellow oil, which was used for subsequent reactions without further purification: δ_H (500 MHz, $CDCl_3$) 8.70 (1H, s, H11), 8.55 (1H, d, *J* 8.5, Ar), 7.63 (1H, dd, *J* 8.5; 1.5, Ar), 7.52 (1H, d, *J* 1.5, Ar), 7.43 (1H, d, *J* 1.5, Ar), 7.24 (1H, dd, *J* 8.2; 1.7, Ar), 6.68 (1H, d, *J* 8.1, Ar), 4.28 (2H, s, H14), 4.20 (2H, t, *J* 6.2, H1 or H4), 4.14 (2H, t, *J* 6.1, H1 or H4), 2.95-2.87 (4H, m, H3 and H6), 2.03-1.92 (4H, m, H2 and H5), 1.57 (9H, s, H8-10), 1.32 (4H, s, H15-16). δ_C (125 MHz, $CDCl_3$) 165.0 (C12 or C13), 165.6 (C12 or C13), 147.2 (Ar), 146.5 (Ar), 141.1 (Ar), 132.8 (Ar), 126.9 (Ar), 124.5 (Ar), 123.8 (Ar), 120.5 (Ar), 118.7 (Ar), 113.7 (Ar), 112.0 (Ar), 111.3 (Ar), 81.3 (C7), 67.3 (C1 or C4), 66.7 (C1 or C4), 39.7 (C3 or C6), 39.6 (C3 or C6), 33.3 (C2 or C5), 33.2 (C2 or C5), 28.7 (C8-10). IR υ_{max} 3432, 3360, 2934, 1703, 1516, 1259. HRMS (ESI) found 459.2602 $C_{24}H_{35}N_4O_5^+$ $[M+H]^+$ requires 459.2602.

***tert*-Butyl 4-(4-amino-3-(3-((*tert*-butoxycarbonyl)amino)propoxy)benzamido)-3-(3-aminopropoxy)benzoate (19)**



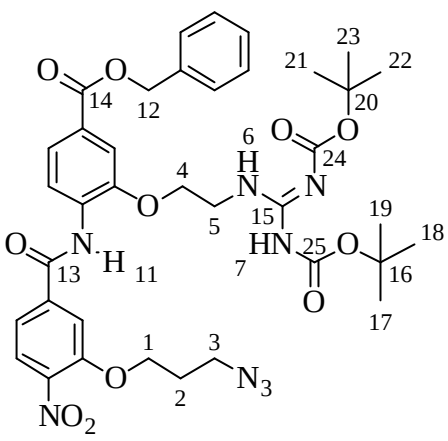
Following the general procedure for catalytic hydrogenation, benzamide **16** (651 mg, 1.06 mmol) was hydrogenated to yield a grey foam (632 mg, 1.13 mmol, *quant*), which was used for subsequent reactions without further purification: δ_H (300 MHz, $CDCl_3$) 8.63 (1H, s,

H12), 8.50 (1H, d, *J* 8.5, Ar), 7.58 (1H, dd, *J* 8.5; 1.7, Ar), 7.47 (1H, d, *J* 1.7, Ar), 7.32 (1H, d, *J* 1.7, Ar), 7.22 (1H, dd, *J* 7.5, 1.7, Ar), 6.64 (1H, d, *J* 8.1, Ar), 5.10 (2H, s, H21), 4.23 (2H, s, H20), 4.14 (2H, t, *J* 6.1, H1 or H4), 4.06 (2H, t, *J* 6.0, H1 or H4), 3.26 (2H, q, *J* 6.1, H3), 2.88 (2H, t, *J* 6.9, H6), 2.00-1.90 (4H, m, H2 and H5), 1.52 (9H, s, H8-10), 1.36 (9H, s, H17-19). δ_C (75 MHz, $CDCl_3$) 165.6 (C11 or C13), 165.1 (C11 or C13), 156.1 (C15), 146.7 (Ar), 145.8 (Ar), 140.7 (Ar), 132.3 (Ar), 126.4 (Ar), 124.0 (Ar), 123.3 (Ar), 120.4 (Ar), 118.2 (Ar), 113.5 (Ar), 111.5 (Ar), 110.6 (Ar), 80.9 (C7), 79.2 (C16), 66.8 (C1 or C4), 66.0 (C1 or C4), 39.2 (C6), 37.6 (C3), 33.0 (C2 or C5), 29.6 (C2 or C5), 28.4 (C17-19), 28.2 (C8-10). IR υ_{max} 3360, 2976, 1697, 1515, 1255, 1165. HRMS (ESI) found 559.3127 $C_{29}H_{43}N_4O_7^+$ $[M+H]^+$ requires 559.3126.

4.10.4. General Procedure for Protected Guanidino Group Formation

DIPEA (2.5 eq) and *N*,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine (1.2 eq) were added to a solution of a primary amine (1 eq) in DCM. The reaction was stirred for 40 hours before water (10 mL/100 mg) and DCM (10 mL/100 mg) were added. The mixture was extracted with DCM (2 x 10 mL/100 mg), dried ($MgSO_4$), filtered, and purified *via* flash column chromatography.

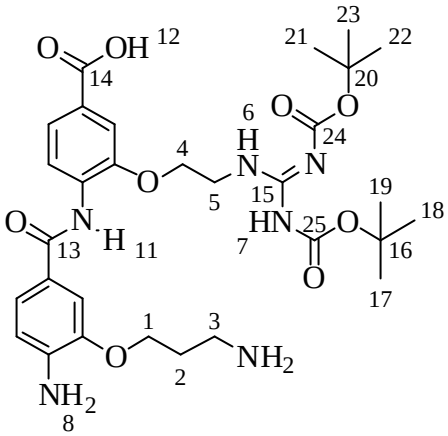
Benzyl-4-(3-(3-azidopropoxy)-4-nitrobenzamido)-3-(2-(2,3-bis(*tert*-butoxycarbonyl) guanidino)ethoxy)benzoate (20)



Following the general procedure for protected guanidino group formation, amine **17** (151 mg, 0.24 mmol) was acylated and purified (3:2 ether:petrol) to yield the title compound as a white solid (131 mg, 0.169 mmol, 71 %): MP 129-131. δ_H (500 MHz,

$CDCl_3$) 11.52 (1H, s, H7), 8.78 (1H, t, *J* 5.5, H6), 8.72 (1H, s, H11), 8.61 (1H, d, *J* 8.5, Ar), 7.96 (1H, d, *J* 8.3, Ar), 7.82 (1H, dd, *J* 8.5; 1.7, Ar), 7.76 (1H, d, *J* 1.5, Ar), 7.62 (1H, d, *J* 1.6, Ar), 7.49-7.34 (6H, m, Ar), 5.37 (2H, s, H12), 4.33-4.27 (4H, m, H1 and H4), 3.95 (2H, q, *J* 5.4, H5), 3.59 (2H, t, *J* 6.3, H3), 2.11 (2H, quint, *J* 6.0, H2), 1.50 (9H, s, H17-19 or H21-23), 1.38 (9H, s, H17-19 or H21-23). δ_C (125 MHz, $CDCl_3$) 166.2 (C14), 163.8 (C13), 163.4 (C15), 156.7 (C24 or C25), 153.9 (C24 or C25), 152.7 (Ar), 147.0 (Ar), 142.1 (Ar), 140.1 (Ar), 136.4 (Ar), 131.8 (Ar), 129.1 (Ar), 128.8 (Ar), 128.7 (Ar), 126.6 (Ar), 126.4 (Ar), 124.4 (Ar), 119.7 (Ar), 118.2 (Ar), 115.0 (Ar), 112.1 (Ar), 84.1 (C16 or C20), 80.2 (C16 or C20), 67.6 (C1), 67.3 (C4), 66.7 (C12), 48.1 (C3), 40.2 (C5), 28.9 (C2), 28.7 (C17-19 or C21-23), 28.3 (C17-19 or C21-23). IR υ_{max} 3421, 3337, 3119, 2978, 2934, 2100, 1717, 1609, 1527, 1272. HRMS (ESI) found 799.3034 $C_{37}H_{44}N_8O_{11}Na^+$ $[M+Na]^+$ requires 799.3022.

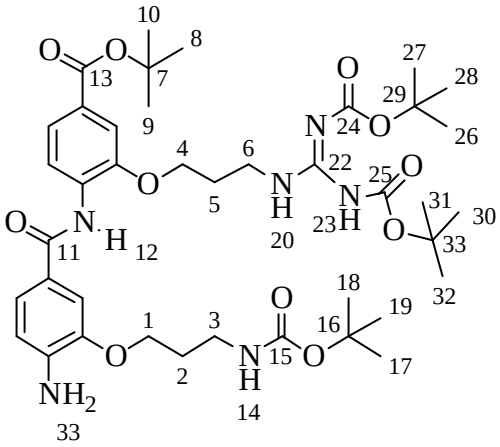
4-(4-Amino-3-(3-aminopropoxy)benzamido)-3-(2-(2,3-bis(*tert*-butoxycarbonyl)guanidino) ethoxy)benzoic acid (21)



Following the general procedure for catalytic hydrogenation, benzamide **20** (118 mg, 0.15 mmol) was hydrogenated to yield the title compound as an off-white powder (60 mg, 0.095 mmol, 63 %), which was used for subsequent reactions without further purification: MP 250-252. δ_H (500 MHz, DMSO) 8.95 (1H, s, H11), 8.62 (1H, t, *J* 6.0, H6), 8.22 (1H, d, *J* 8.3, Ar), 7.58 (1H, d, *J* 8.4, Ar), 7.56 (1H, s, Ar), 7.37 (1H, d, *J* 1.5, Ar), 7.33 (1H, dd, *J* 8.3; 1.6, Ar), 6.68 (1H, d, *J* 8.2, Ar), 5.60 (2H, s, H8), 4.25 (2H, t, *J* 5.2, H4), 4.12 (2H, t, *J* 5.9, H1), 3.80 (1H, q, *J* 5.1, H5), 3.04 (2H, t, *J* 7.1, H3), 2.06 (2H, t, *J* 6.7, H2), 1.39 (18H, s, H17-19 and H21-23). δ_C (125 MHz,

DMSO) 164.7 (C13 or C14), 164.6 (C13 or C14), 163.0 (C15), 155.5 (C24 and C25), 152.0 (Ar), 147.4 (Ar), 145.1 (Ar), 144.5 (Ar), 142.2 (Ar), 131.2 (Ar), 122.4 (Ar), 121.1 (Ar), 120.7 (Ar), 119.9 (Ar), 112.0 (Ar), 110.7 (Ar), 83.0 (C16 or C20), 78.3 (C16 or C20), 66.3 (C1 or C4), 65.0 (C1 or C4), 36.4 (C3), 27.9 (C2), 27.6 (C17-19 or C21-23), 27.5 (C17-19 or C21-23). IR υ_{max} 3331, 2928, 1718, 1613, 1517, 1254. HRMS (ESI) found 631.3082 $C_{30}H_{43}N_6O_9^+$ $[M+H]^+$ requires 631.3086.

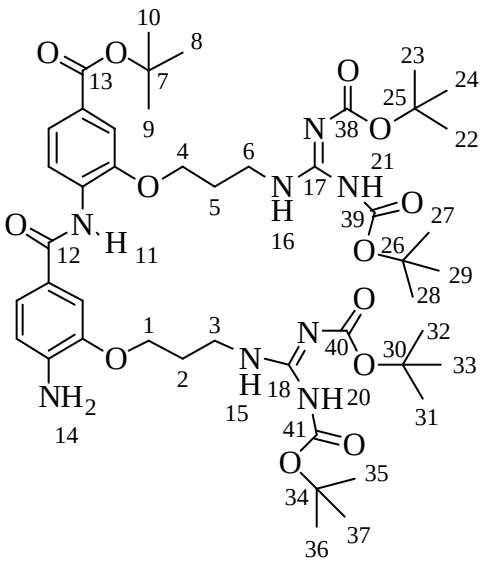
***tert*-Butyl 4-(4-amino-3-(3-((*tert*-butoxycarbonyl)amino)propoxy)benzamido)-3-(3-(2,3-bis(*tert*-butoxycarbonyl)guanidino)propoxy)benzoate (22)**



Following the general procedure for protected guanidino group formation, amine **19** (592 mg, 1.06 mmol) was reacted for 5 days and purified (3:1 ether:petrol) to yield the title compound (767 mg, 0.96 mmol, 90 %) as a white foam: δ_H (500 MHz, $CDCl_3$) 11.46 (1H, s, H23), 8.65 (1H, s, H12), 8.52-8.47 (2H, m, Ar and H20), 7.62 (1H, d, *J* 7.8, Ar), 7.49 (1H, s, Ar), 7.43 (1H, s, Ar), 7.28 (1H, dd, *J* 8.2; 0.9, Ar), 6.68 (1H, d, *J* 8.2, Ar), 4.92 (1H, s, H14), 4.29 (2H, s, H33), 4.16 (2H, t, *J* 5.8, H4), 4.09 (2H, t, *J* 5.9, H1), 3.66 (2H, q, *J* 6.1, H6), 3.30 (2H, q, *J* 5.8, H3), 2.12 (2H, quint, *J* 6.1, H5), 1.97 (2H, quint, *J* 6.1, H2), 1.56 (9H, s, H 8-10), 1.39 (27H, s, H17-19 and H26-28 and H30-32). δ_C (125 MHz, $CDCl_3$) 166.0 (C11 or C13), 165.5 (C11 or C13), 163.9 (C22), 156.8 (C15 or C24 or C25), 156.5 (C15 or C24 or C25), 153.7 (C15 or C24 or C25), 147.2 (Ar), 146.2 (Ar), 141.1 (Ar), 132.6 (Ar), 127.0 (Ar), 124.3 (Ar), 123.8 (Ar), 120.9 (Ar), 119.3 (Ar), 113.9 (Ar), 111.9 (Ar), 111.3 (Ar), 83.8 (C16 or C29 or C33), 81.3 (C7), 79.8 (C16 or C29 or C33), 79.6 (C16 or C29 or C33), 66.7 (C4), 66.4 (C1), 39.4 (C6), 38.2 (C3), 30.0 (C2), 29.4 (C5),

28.8 (C8-10 or C17-19 or C26-28 or C30-32), 28.6 (C8-10 or C17-19 or C26-28 or C30-32), 28.4 (C8-10 or C17-19 or C26-28 or C30-32). IR υ_{max} 3342, 2977, 1710, 1617, 1515, 1273, 1165. HRMS (ESI) found 823.4224 $C_{40}H_{60}N_6O_{11}Na^+$ $[M+Na]^+$ requires 823.4212.

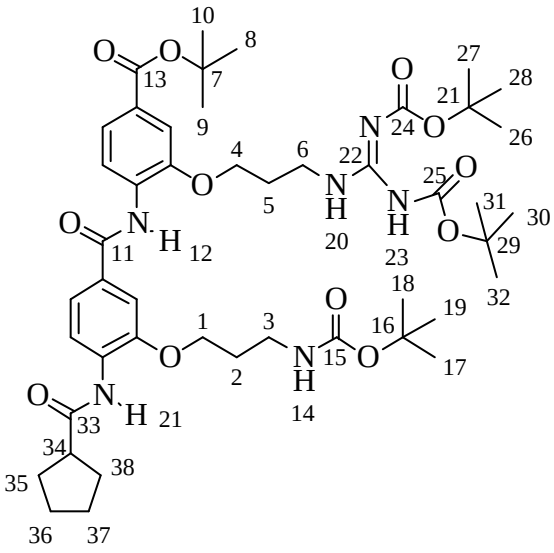
***tert*-Butyl 4-(4-amino-3-(3-(2,3-bis(*tert*-butoxycarbonyl)guanidino)propoxy) benzamido)-3-(3-(2,3-bis(*tert*-butoxycarbonyl)guanidino)propoxy)benzoate (23)**



Following the general procedure for protected guanidino group formation di-amine **18** (165 mg, 0.36 mmol, 1.0 eq) was reacted with DIPEA (185 mg, 1.43 mmol, 4.0 eq) and *N*,*N*'-di-Boc-1*H*-pyrazole-1-carboxamidine (255 mg, 0.82 mmol, 2.3 eq) and purified (2:1 ether:petrol) to yield the title compound (261 mg, 0.28 mmol, 77 %) as a colourless oil: δ_H (500 MHz, $CDCl_3$) 11.51 (1H, s, H20 or H21), 11.47 (1H, s, H20 or H21), 8.66 (1H, s, H11), 8.51 (1H, t, *J* 5.0, H15 or H16), 8.46 (1H, t, *J* 5.0, H16 or H15), 7.64 (1H, dd, *J* 8.5; 1.6, Ar), 7.57 (1H, d, *J* 1.9, Ar), 7.50 (1H, d, *J* 1.5, Ar), 7.48 (1H, d, *J* 1.5, Ar), 7.28 (1H, dd, *J* 8.2; 1.6, Ar), 6.69 (1H, d, *J* 8.2, Ar), 4.33 (2H, s, H14), 4.21-4.13 (4H, m, H1 and H4), 3.72-3.61 (4H, m, H3 and H6), 2.18-2.06 (4H, m, H2 and H5), 1.57 (9H, s, H8-10), 1.48 (H22-24 or H27-29 or H31-33, or H35-37), 1.47 (H22-24 or H27-29 or H31-33 or H35-37), 1.42 (H22-24 or H27-29 or H31-33 or H35-37), 1.41 (H22-24 or H27-29 or H31-33 or H35-37). δ_C (125 MHz, $CDCl_3$) 166.0 (C12 or C13), 165.5 (C12 or C13), 164.0 (C17 and C18), 156.8 (C38 or C39 or C40 or C41), 156.6 (C38 or C39 or C40 or C41), 153.7 (C38 or C39 or C40 or C41), 153.7 (C38 or C39 or C40 or C41), 147.2 (Ar), 146.2 (Ar), 141.1 (Ar), 132.7 (Ar), 127.0 (Ar), 124.3 (Ar), 123.8 (Ar), 120.7 (Ar), 119.3 (Ar), 114.0 (Ar), 111.9 (Ar), 111.5 (Ar), 83.7 (C25 or C26 or

C30 or C34), 83.6 (C25 or C26 or C30 or C34), 81.3 (C7), 79.8 (C25 or C26 or C30 or C34), 79.7 (C25 or C26 or C30 or C34), 66.8 (C1 or C4), 66.7 (C1 or C4), 39.0 (C3 or C6), 38.5 (C3 or C6), 29.4 (C2 or C5), 29.2 (C2 or C5), 28.7 (C22-24 or C27-29 or C31-33 or C35-37), 28.7 (C22-24 or C27-29 or C31-33 or C35-37 and C8-10), 28.5 (C22-24 or C27-29 or C31-33 or C35-37), 28.4 (C22-24 or C27-29 or C31-33 or C35-37). IR υ_{max} 3329, 2979, 1718, 1614, 1327, 1131. HRMS (ESI) found 965.4964 $C_{46}H_{70}N_8O_{13}Na^+$ $[M+Na]^+$ requires 965.4955.

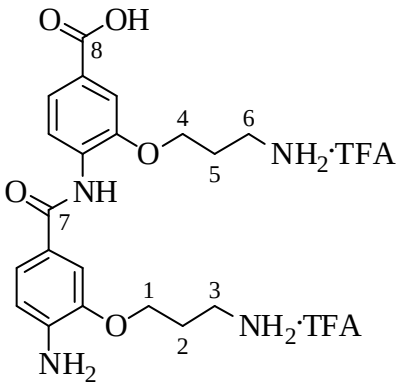
tert-Butyl 3-(3-(2,3-bis(_tert_-butoxycarbonyl)guanidino)propoxy)-4-(3-(3-((_tert_-butoxycarbonyl)amino)propoxy)-4-cyclopentanecarboxamido)benzamido)benzoate (24)



Following the general procedure for amidation, cyclopentanecarboxylic acid (16 mg, 0.14 mmol, 1.1 eq) was coupled with amine **22** (102 mg, 0.13 mmol, 1.0 eq) and purified (3:2 ether:petrol) to yield the title compound (78 mg, 0.087 mmol, 69 %) as an off-white wax: δ_H (500 MHz, $CDCl_3$) 11.49 (1H, s, H23), 8.77 (1H, s, H12), 8.53 (1H, s, H20), 8.52-8.48 (2H, m, Ar), 8.37 (1H, s, H21), 7.66 (1H, dd, *J* 8.4; 1.5, Ar), 7.58 (1H, d, *J* 1.6, Ar), 7.52 (1H, d, *J* 1.4, Ar), 7.48 (1H, dd, *J* 8.5; 1.5, Ar), 4.72 (1H, s, H14), 4.22-4.17 (4H, m, H1 and H4), 3.67 (2H, q, *J* 6.2, H6), 3.36 (2H, q, *J* 6.3, C3), 2.92 (1H, quint, *J* 7.8, H34), 2.15 (2H, quint, *J* 6.3, H5), 2.05-1.96 (4H, m, H2 and H35 or H38), 1.93-1.87 (2H, m, H35 or H38), 1.83-1.76 (2H, m, H36 and H37), 1.66-1.60 (2H, m, H36 and H37), 1.59 (9H, s, H8-10), 1.44 (9H, s, H17-19 or H26-28 or H30-32), 1.42 (9H, s, H17-19 or H26-28 or H30-32), 1.41

(9H, s, H17-19 or H26-28 or H30-32). δ_C (125 MHz, $CDCl_3$) 175.7 (C33), 165.9 (C11 or C13), 165.2 (C11 or C13), 164.0 (C22), 156.8 (C15 or C24 or C25), 156.5 (C15 or C24 or C25), 153.7 (C15 or C24 or C25), 147.5 (Ar), 147.4 (Ar), 132.4 (Ar), 132.2 (Ar), 129.8 (Ar), 127.5 (Ar), 123.7 (Ar), 120.2 (Ar), 119.8 (Ar), 119.6 (Ar), 112.0 (Ar), 111.0 (Ar), 83.7 (C16 or C21 or C29), 81.4 (C7), 79.9 (C16 or C21 or C29), 79.7 (C16 or C21 or C29), 66.5 (C4), 65.8 (C1), 47.3 (C34), 38.2 (C6), 37.4 (C3), 31.0 (C35 and C38), 30.4 (C2), 29.4 (C5), 28.8 (C8-10 or C17-19 or C26-28 or C30-32), 28.7 (C8-10 or C17-19 or C26-28 or C30-32), 28.6 (C8-10 or C17-19 or C26-28 or C30-32), 28.4 (C8-10 or C17-19 or C26-28 or C30-32), 26.5 (C36-37). IR υ_{max} 3425, 3326, 2975, 1710, 1641, 1517, 1166. HRMS (ESI) found 919.4793 $C_{46}H_{68}N_6O_{12}Na^+$ $[M+Na]^+$ requires 919.4787.

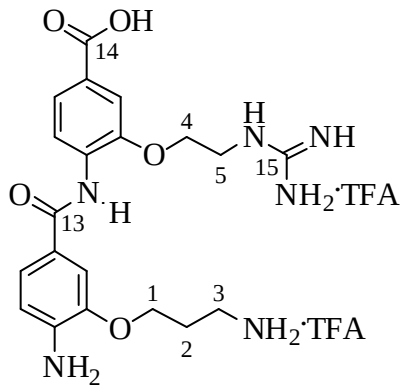
4-(4-Amino-3-(3-aminopropoxy)benzamido)-3-(3-aminopropoxy)benzoic acid•2TFA (1)



Following the general procedure for *N*-Boc deprotection, protected benzamide **18** (150 mg, 0.327 mmol) was stirred for 30 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (25 % MeCN to 75 % MeCN over 40 minutes) to yield the title compound (6 mg, 9.5 μmol) as a yellow oil: δ_H (400 MHz, D_2O) 7.68 (1H, d, *J* 8.3, Ar), 7.54 (1H, d, *J* 8.3, Ar), 7.47 (1H, s, Ar), 7.28-7.23 (2H, m, Ar), 6.89 (1H, d, *J* 8.7, Ar), 4.16-4.08 (4H, m, H1 and H4), 3.18 (2H, t, *J* 7.4, H3 or H6), 3.07 (2H, t, *J* 7.4, H3 or H6), 2.18-2.06 (4H, m, H2 and H5). δ_C (100 MHz, D_2O) 170.2 (C7 or C8), 168.1 (C7 or C8), 163.0 (q, *J* 35.5, CF_3CO_2), 149.9 (Ar), 146.6 (Ar), 138.7 (Ar), 130.4 (Ar), 128.3 (Ar), 124.4 (Ar), 123.5 (Ar), 123.0 (Ar), 121.4 (Ar), 116.3 (q, *J* 291.6, CF_3CO_2), 116.1 (Ar), 113.2 (Ar), 111.2 (Ar), 65.8 (C1 or C4), 65.7 (C1 or C4), 37.2 (C3 or C6), 37.0 (C3 or C6),

26.5 (C2 or C5), 26.4 (C2 or C5). IR υ_{max} 3359, 3067, 2959, 1677, 1520, 1201, 1135. HRMS (ESI) found 403.1971 $C_{20}H_{27}N_5O_5^+$ $[M+H]^+$ requires 403.1976.

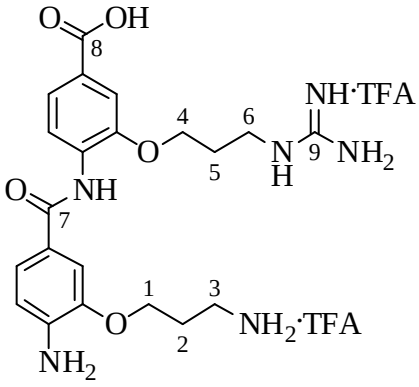
1-(2-(2-(4-Amino-3-(3-ammoniopropoxy)benzamido)-5-carboxyphenoxy)ethyl)guanidinium•2 TFA (2)



Following the general procedure for *N*-Boc deprotection, protected benzamide **21** (38 mg, 0.06 mmol) was stirred for 90 minutes and the residue purified *via* semi-preparative HPLC (10 % MeCN-35 % MeCN over 40 minutes) to yield the title compound as a yellow oil (6.8 mg, 0.01 mmol): δ_H (500 MHz, D_2O) 7.72 (1H, d, *J* 8.3, Ar), 7.65 (1H, dd, *J* 3.3; 1.7, Ar), 7.61 (1H, d, *J* 1.6, Ar), 7.49 (1H, d, *J* 2.0, Ar), 7.45 (1H, dd, *J* 8.2; 1.7, Ar), 7.37 (1H, d, *J* 8.2, Ar), 4.26-4.21 (4H, m, H1 and H4), 3.55 (2H, t, *J* 4.9, H5), 3.19 (2H, t, *J* 7.5, H3), 2.18 (2H, quint, *J* 6.3, H2). δ_C (125 MHz, D_2O) 169.6 (C13 or C14), 167.8 (C13 or C14), 162.9 (q, *J* 36, F_3CCO_2), 157.1 (C15), 150.7 (Ar), 159.4 (Ar), 133.5 (Ar), 130.4 (Ar), 128.2 (Ar), 125.6 (Ar), 124.4 (Ar), 123.4 (Ar), 122.8 (Ar), 120.6 (Ar), 116.3 (q, *J* 292, F_3CCO_2), 113.9 (Ar), 111.9 (Ar), 67.0 (C1 or C4), 66.1 (C1 or C4), 40.8 (C5), 36.9 (C3), 26.4 (C2). IR υ_{max} 3358, 3177, 2951, 1675, 1518, 1201, 1135. HRMS (ESI) found 431.2037 $C_{20}H_{27}N_6O_5^+$ $[M+H]^+$ requires 431.2037.

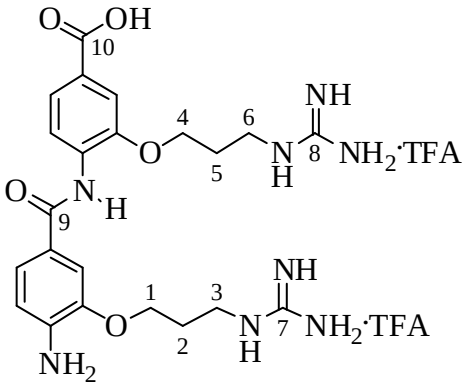
4-(4-Amino-3-(3-aminopropoxy)benzamido)-3-(3-guanidinopropoxy)benzoic acid•2TFA (3)

Following the general procedure for *N*-Boc deprotection, protected benzamide **18** (60 mg, 0.075 mmol) was stirred for 80 minutes in TFA. A portion of the residue was purified *via* semi-preparative HPLC (15 % MeCN to 45 % MeCN over 40 minutes) to yield the title



compound (10 mg, 0.015 mmol) as a yellow oil: δ_H (500 MHz, D_2O) 7.80 (1H, d, *J* 8.3, Ar), 7.56 (1H, d, *J* 8.2, Ar), 7.48 (1H, d, *J* 6.7, Ar), 7.46-7.37 (3H, m, Ar), 4.25 (2H, t, *J* 5.6, H1 or H4), 4.14-4.08 (2H, m, H1 or H4), 3.30 (2H, t, *J* 6.7, H3 or H6), 3.26 (2H, t, *J* 7.5, H3 or H6), 2.24 (2H, quint, *J* 7.8, H2 or H5), 2.04 (2H, quint, *J* 6.3, H2 or H5). δ_C (125 MHz, D_2O) 169.4 (C7 or C8), 166.9 (C7 or C8), 162.8 (q, *J* 35.6, CF_3CO_2), 156.7 (C9), 151.0 (Ar), 149.7 (Ar), 134.1 (Ar), 130.3 (Ar), 127.3 (Ar), 124.4 (Ar), 123.3 (Ar), 122.9 (Ar), 122.8 (Ar), 120.2 (Ar), 116.2 (q, *J* 291.7, CF_3CO_2), 113.0 (Ar), 111.6 (Ar), 66.0 (C1 or C4), 65.9 (C1 or C4), 38.1 (C3 or C6), 36.9 (C3 or C6), 27.6 (C2 or C5), 26.4 (C2 or C5). IR υ_{max} 3361, 3172, 2959, 1674, 1603, 1198, 1136. HRMS (ESI) found 445.2193 $C_{21}H_{29}N_6O_5^+$ $[M+H]^+$ requires 445.2194.

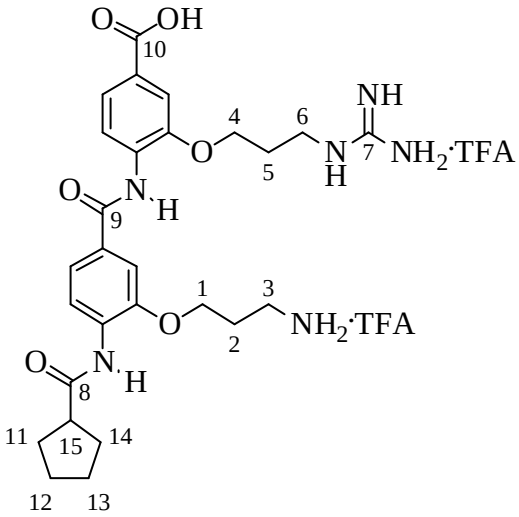
4-(4-amino-3-(3-guanidinopropoxy)benzamido)-3-(3-guanidinopropoxy)benzoic acid•2 TFA (4)



Following the general procedure for *N*-Boc deprotection, protected benzamide **23** (93 mg, 0.10 mmol) was stirred for 105 minutes and the residue purified *via* semi-preparative HPLC (15 % MeCN-45 % MeCN over 40 minutes) to yield the title compound as a yellow oil (11.1 mg, 0.016 mmol): δ_H (500 MHz, D_2O) 7.71 (1H, d, *J* 8.3, Ar), 7.59 (1H, dd, *J* 8.3; 1.7, Ar), 7.55 (1H, d, *J* 1.6, Ar), 7.45 (1H, d, *J* 1.3, Ar), 7.42 (1H, dd, *J* 8.2; 1.6, Ar), 7.37 (1H, d, *J* 8.2, Ar), 4.17 (2H, t, *J* 6.0, H1 or H4), 4.13 (2H, t, *J* 5.9, H1 or H4), 3.34 (2H, t, *J* 6.8, H3 or H6), 3.24 (2H, t, *J* 6.7, H3 or H6), 2.07 (2H, quint, *J* 6.4, H2 or H5), 1.99 (2H, quint, *J* 6.3, H2 or H5). δ_C (125

MHz, D_2O) 169.6 (C9 or C10), 167.7 (C9 or C10), 162.9 (q, *J* 36, CF_3CO_2), 156.8 (C7 or C8), 156.7 (C7 or C8), 151.0 (Ar), 150.5 (Ar), 130.8 (Ar), 130.3 (Ar), 128.0 (Ar), 125.0 (Ar), 124.0 (Ar), 123.0 (Ar), 123.0 (Ar), 120.3 (Ar), 116.3 (q, *J* 292, CF_3CO_2), 113.5 (Ar), 111.7 (Ar), 66.0 (C1 or C4), 65.9 (C1 or C4), 38.1 (C3 or C6), 37.9 (C3 or C6), 27.6 (C2 or C5), 27.5 (C2 or C5). IR υ_{max} 3350, 3194, 2957, 1670, 1604, 1199, 1137. HRMS (ESI) found 487.2397 $C_{22}H_{31}N_8O_5^+$ $[M+H]^+$ requires 487.2412.

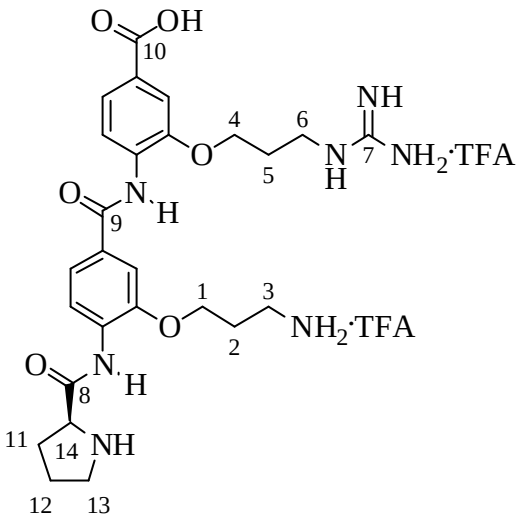
4-(3-(3-Aminopropoxy)-4-(cyclopentanecarboxamido)benzamido)-3-(3-guanidinopropoxy)benzoic acid•2TFA (5)



Following the general procedure for *N*-Boc deprotection, benzamide **24** (76 mg, 0.085 mmol) was reacted for 2 hours and the residue purified *via* semi-preparative HPLC (25 %-50 % MeCN) to yield the title compound (12.9 mg, 0.018 mmol) as a white solid: MP 93-95. δ_H (500 MHz, D_2O) 7.72 (1H, d, *J* 8.3, Ar), 7.60 (1H, d, *J* 8.3, Ar), 7.31 (1H, dd, *J* 8.3; 1.5, Ar), 7.07 (1H, d, *J* 0.8, Ar), 6.91 (1H, s, Ar), 6.81 (1H, d, *J* 8.2, Ar), 3.91-3.85 (4H, m, H1 and H4), 3.27 (1H, t, *J* 6.9, H3 or H6), 3.09 (2H, t, *J* 7.7, H3 or H6), 2.79-2.71 (1H, m, H15), 2.08 (2H, quint, *J* 6.3, H2 or H5), 1.97 (2H, quint, *J* 6.5, H2 or H5), 1.91-1.84 (2H, m, H11 and H14), 1.68-1.53 (6H, m, H11-14). δ_C (125 MHz, D_2O) 178.3 (C8), 169.2 (C9 or C10), 165.0 (C9 or C10), 163.0 (q, *J* 36, CF_3CO_2), 156.8 (C7), 148.4 (Ar), 147.5 (Ar), 130.4 (Ar), 129.9 (Ar), 128.8 (Ar), 125.5 (Ar), 122.8 (Ar), 121.1 (Ar), 119.7 (Ar), 118.8 (Ar), 116.3 (q, *J* 292, CF_3CO_2), 111.7 (Ar), 110.3 (Ar), 65.6 (C1 or C4), 65.4 (C1 or C4), 46.1 (C15), 38.2 (C3 or C6), 36.9 (C3 or C6), 30.2 (C11 and

14), 27.7 (C2 or C5), 26.5 (C2 or C5), 25.6 (C12-13). IR υ_{max} 3412, 3367, 2960, 1667, 1519, 1198, 1134. HRMS (ESI) found 541.2760 $C_{27}H_{37}N_6O_6^+$ $[M+H]^+$ requires 541.2769.

(*S*)-4-(3-(3-Aminopropoxy)-4-(pyrrolidine-2-carboxamido)benzamido)-3-(3-guanidinopropoxy)benzoic acid•2 TFA (6)



Following the general procedure for amidation, *N*-Boc-L-proline (22 mg, 0.1 mmol) was coupled with amine **22** (86 mg, 0.105 mmol). After workup, following the general procedure for *N*-Boc deprotection, the protected molecule was reacted for 100 minutes and the residue purified *via* semi-preparative HPLC (15 %-45 % MeCN over 40 minutes) to yield the title compound (18.5 mg, 0.025 mmol) as a colorless oil: $[\alpha]_D^{25}$ +17.5 (T=24 ºC, c=1.6, MeOH). δ_H (500 MHz, D_2O) 7.74 (1H, d, *J* 8.4, Ar), 7.69 (1H, d, *J* 8.2, Ar), 7.43-7.39 (1H, m, Ar), 7.28 (1H, s, Ar), 7.20-7.14 (2H, m, Ar), 4.59 (1H, dd, *J* 8.7; 6.7, H14), 4.09-3.97 (4H, m, H1 and H4), 3.49-3.36 (2H, m, H13), 3.25 (2H, t, *J* 6.8, H3 or H6), 3.14 (2H, t, *J* 7.6, H3 or H6), 2.56-2.47 (1H, m, H11), 2.19-2.10 (3H, m, H2 or H5 and H11), 2.09-2.02 (2H, m, H12), 2.02-1.94 (2H, m, H2 or H5). δ_C (125 MHz, D_2O) 169.4 (C8 or C9 or C10), 169.3 (C8 or C9 or C10), 168.2 (C8 or C9 or C10), 163.0 (q, *J* 36, $CF_3\mathit{C}O_2$), 156.8 (C7), 149.6 (Ar), 148.8 (Ar), 130.5 (Ar), 130.4 (Ar), 128.8 (Ar), 123.1 (Ar), 122.9 (Ar), 122.6 (Ar), 121.5 (Ar), 119.8 (Ar), 116.3 (q, *J* 292, $\mathit{C}F_3CO_2$), 112.6 (Ar), 110.9 (Ar), 65.9 (C1 or C4), 65.8 (C1 or C4), 60.3 (C14), 46.7 (C13), 38.2 (C3 or C6), 36.9 (C3 or C6), 30.1 (C11), 27.6 (C2 or C5), 26.4 (C2 or C5), 23.8 (C12). IR υ_{max} 3351, 3179, 2970, 1673, 1526, 1201, 1131. HRMS (ESI) found 542.2728 $C_{26}H_{36}N_7O_6^+$ $[M+H]^+$ requires 542.2722.

5. Thermodynamics of hydrogen bond-controlled molecular switches by DFT

5.1. Introduction

In the systems described in Chapters 3 & 4, the molecules were held in relatively well defined conformations by single hydrogen bonds. These intramolecular hydrogen bonds were correctly calculated even with unsophisticated computational methods, and were validated by NMR and crystallographic measurements. In contrast to these largely restricted systems, other molecules of organic or biological relevance rely on much more complicated hydrogen bonding. On a large scale, proteins form interconnected H-bond networks between backbone atoms, sidechains, and water.[1] These networks are exploited by antifreeze proteins, which bond to nascent ice crystals to depress the freezing point of water and permit organisms to survive at low temperatures.[2] Multiple hydrogen bonds are also a feature of numerous molecular recognition systems, some of which bind with sufficient affinity to operate even in aqueous media.[3] Furthermore, many asymmetric catalysts use multiple hydrogen bonds to direct the formation of specific substrate-catalyst complexes.[4] The accurate, quantitative prediction of interactions in such systems would be of great benefit to organic chemists in designing future molecules, but this has proved challenging.

Much recent computational work has highlighted the importance of accurately accounting for non-bonding interactions in the development of electronic structure theory.[5–7] While high-level computational methods do account for non-covalent interactions such as dispersion and torsional parameters, such techniques are far too computationally expensive to be applied to systems with more than a few atoms. Instead, less-precise density functional theory (DFT) methods are used by the great majority of organic chemists to model compounds of interest. In order to improve these necessarily imperfect calculations, it is critical to assess the types of systems for which current DFT methods fail. Large

benchmarking datasets at the gold-standard[8] CCSD(T)/CBS level of theory, including the S22 and S66 sets of Hobza and co-workers for small complexes bound by hydrogen-bonding and dispersion interactions,[9,10] the composite GMTKN30 set of Goerigk and Grimme,[11] and a large composite database constructed by Friesner and co-workers,[12] provide clear standards against which new methods can be developed and evaluated. Using these and other benchmarking sets, improved descriptions of non-bonding interactions have been implemented into semi-empirical DFT methods, such as hybrid meta-GGA functionals (e.g. the Minnesota family),[13] explicit dispersion corrections for London dispersion such as D3,[14] and *van der Waals* functionals.[15,16] Much specific work has been directed towards understanding hydrogen bonding, with recent reports investigating DFT's ability to reproduce the distances[17] and angles[18] of small H-bonded complexes. While forcefields systematically underestimate the stability of such complexes,[19] accurate potential energy curves can be calculated with appropriately-chosen DFT methods.[20,21]

The Hamilton group recently reported the synthesis and characterization of a series of benzamido-diphenylacetylene (DPA) molecular switches in which the relative hydrogen bonding capability of two amide groups is the critical determinant of conformation.[22–24] In this system, two relatively planar, hydrogen-bonded conformations can readily interconvert by a 180º rotation about the acetylene axis. Different substituents affect the electronics of the amides and thus change the relative strengths of the two intramolecular hydrogen bonds. By varying the attached functionalities (at positions R_1 and R_2) the conformational preference of the system is changed (Figure 5-1).

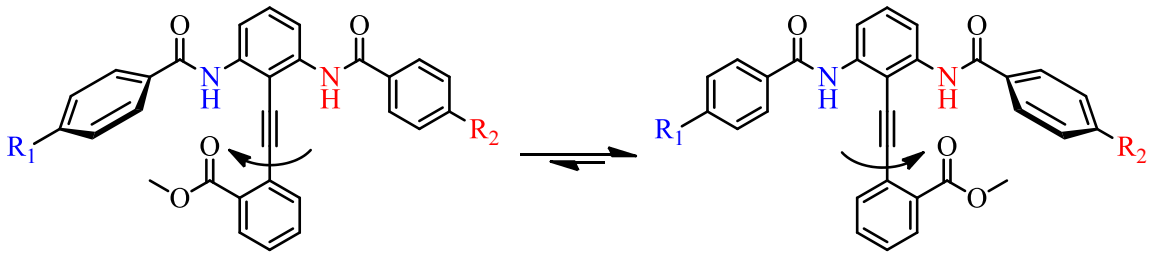


Figure 5-1: Two rotamers of the DPA-based molecular switch defined by a 180º rotation about the DPA axis. By varying the two substituents, the relative acidities of the amides is changed, allowing control over the equilibrium conformation of the molecule. Due to steric clashing with the methyl ester, the proximal benzamide ring rotates out of the plane.

Ten molecules of this class have been synthesized, incorporating electron withdrawing groups (*para*-nitro **5**, *para*-chloro **3**, 4-pyridyl **8**), electron donating groups (*para*-methoxy **7**, *para*-dimethylamino **6**), mixed cases (*para*-methoxy on one benzamide and *para*-chloro **2** or *para*-nitro **4** on the other), as well as 2-pyridyl substitution **9** and molecule **10** which substitutes a urea in place of one of the benzamides (Table 5-1).

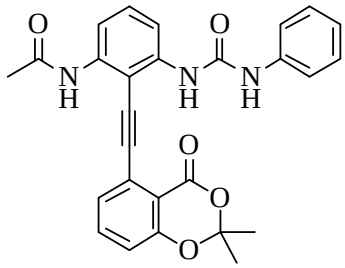
The equilibrium conformational ratio is measured by comparing the NMR shifts of the amide NHs in chloroform with those of control molecules which are incapable of forming hydrogen bonds.[22–24] Using this assay, the equilibria of ten molecules have been experimentally determined, and were found to encompass relative hydrogen bond strengths spanning a range of approximately 2 kcal.mol^{-1}, with most of the molecules clustered in a 1 kcal.mol^{-1} range. Free energy differences between the "on" and "off" forms of the switches are calculated using the definition of Gibbs free energy:

$$\Delta G = -RT\ln(K)$$

where R is the gas constant, T is the experimental temperature of 298.15 K, and K is the equilibrium conformational ratio determined by the NMR assay.

Table 5-1: Subsituents and experimental free energies of the DPA molecular switches. Molecules marked with a star are not substituents on a phenyl ring, but rather represent replacement of the ring with the marked group (e.g. molecule 8 juxtaposes a *para*-pyridine ring and a phenyl ring). Molecule **10** is of a different class, and features acetamide and phenylurea rather than the benzamides of the other molecules.

Molecule	R^1	R^2	ΔG (kcal.mol^{-1})
1	H	H	0.00
2	Cl	MeO	-0.21
3	Cl	H	-0.17
4	NO_2	MeO	-0.44
5	NO_2	H	-0.37
6	NMe_2	H	0.17
7	MeO	H	0.03
8	4-pyridyl*	H	-0.38
9	2-pyridyl*	H	0.63
10	NHPh*	Me*	-1.37



10

These molecular switches provide a difficult test for computation. They are too large to be optimized with correlated *ab initio* methods, and so DFT must be used. Furthermore, the conformation is dictated by the relative magnitude of hydrogen bonding interactions in concert with torsional rotational potentials and other non-bonding contacts. The ability of computational methods to accurately account for the observed behavior spanning only a few kcal.mol^{-1} is of importance, since functional group effects of similar magnitude are fundamental to organic chemistry. Given the precision of the NMR assay in measuring the equilibrium constants for the DPA switches, this system serves as a novel experimental dataset. Moreover, although high level *ab initio* calculations provide the "gold standard" of computation, it is useful to compare to experiment whenever possible.

5.2. Computational methods

Structures (of both rotameric forms) were built using *Gaussview 5*[25]. All Hartree-Fock (HF) and DFT calculations were performed with *Gaussian09*, rev B.01.[26] Optimizations were performed using HF, the B3LYP hybrid generalized gradient approximation (GGA) functional,[27,28] the long-range corrected CAM-B3LYP which increases the amount of HF

exchange at long-range,[29] the M06-2X[13] hybrid meta-GGA functional, and the dispersion-corrected ωB97X-D[30] functional in combination with a Pople 6-31+G(d,p) basis set for all elements.[31–33] Standard convergence criteria were used along with a fine grid for all numerical integrations. Stationary points on the potential energy surface were confirmed as minima by the presence of no imaginary harmonic vibrational frequencies. Free energies include unscaled zero point vibrational energies. Where indicated, low frequencies (< 100 cm^{-1}) are corrected in the vibrational component of the entropy using a rigid rotor approximation according to the method of Grimme *et al*, since entropy associated with these loose vibrational modes is the most prone to computational error.[34] Optimizations were performed in the gas phase or with the implicit conductor-like polarizable continuum model (CPCM) of the NMR solvent, chloroform.[35] Single point calculations were performed at stationary points using Dunning's quadruple-ζ cc-pVQZ basis set.[36] DFT methods in general converge quickly to the Kohn-Sham limit,[37] although not necessarily in a monotonic way.[38,39] To verify that the cc-pVQZ is sufficiently large to assure basis set convergence, extrapolation using the two-point formula of Halkier *et. al.*[40] was conducted:

$$E_{CBS} \approx \frac{E_X X^3 - E_Y Y^3}{X^3 - Y^3}$$

where X and Y are the cardinal numbers of the Dunning basis sets used for the extrapolation. Based on extrapolations from cc-pVTZ and cc-pVQZ optimizations in solvent using B3LYP and ωB97X-D, basis sets beyond the quadruple-ζ level would change the calculated energy differences of the switches by < 0.02 $kcal.mol^{-1}$. Where indicated, dispersion correction to energies were applied using the DFT-D3 correction with zero-damping and the three-body term included.[14] Unless otherwise stated, all energies are quoted in $kcal.mol^{-1}$ and Gibbs free energies (incorporating ZPE, thermal, and entropic contributions obtained with the 6-31+G(d,p) basis set) are at 298.15 K and a standard state of 1 $mol.L^{-1}$.

5.2.1. Statistical analysis

No single statistical parameter is sufficient to assess the accuracy and utility of the computational methods. For instance, error statistics such as mean unsigned errors (MUEs) and root-mean-square deviations (RMSDs) describe the absolute energy differences between calculation and experiment, but may hide the utility of methods which accurately model the trend in relative energies. As the key question in molecular design is often "which molecule will show the greater effect?" and not "what will be the absolute effect size?" a method which correctly ranks the compounds would be useful even if the absolute energies are not modeled correctly. As a result, the agreement between predicted and experimental free energy differences of the rotamers was assessed using MUEs, RMSDs, linear coefficient of determination (r^2), and Pearlman and Charifson's predictive index (PI).[41,42] The latter quantifies the ability of the calculations to appropriately order the molecules in energy difference, with a value of +1 indicating perfect ranking and -1 indicating perfectly anticorrelated predictions. Due to the very limited energy range of the system, a calculation that predicted all of the switches to have no bias ($\Delta G = 0$) would have an RMSD of 0.56 kcal.mol^{-1} and an MUE of 0.4 kcal.mol^{-1}. Despite these low errors, such a calculation would be useless in ranking the compounds (PI undefined) and would show no linear correlation with the data.

5.3. Optimizations

5.3.1. Gas phase

In the first instance, the structures were optimized in the gas phase and energies determined by a single point calculation in CPCM chloroform. Comparison with the experimental data indicates relatively poor performance for several of the DFT methods (Table 5-2). In particular, while B3LYP and the related CAM-B3LYP showed the lowest

errors and greatest correlation with experiment, the more recently developed ωB97X-D and M06-2X functionals fared substantially worse, with errors more than double those of B3LYP. The D3 dispersion correction of Grimme added to the B3LYP geometries had the overall effect of increasing the calculated energy differences, resulting in higher errors but better PI. HF had errors nearly as low as calculated by B3LYP, but the low coefficient of determination and PI reveal that HF cannot reliably distinguish between the molecules; the low errors are a serendipitous result of the small data range and calculations predicting all molecules to be roughly equal (Figure 5-2).

Table 5-2: List of statistical results from benchmarking calculations in the gas phase using HF and several hybrid DFT functionals. The structures of each rotamer were optimized in the gas phase with the listed functional and the 6-31+G(d,p) basis set. Electronic energies were calculated from a single point calculation in CPCM chloroform using the same functional and the cc-pVQZ basis set. Listed are the statistical parameters comparing calculation to experiment.

Functional	MUE	RMSD	r^2	PI
HF	0.39	0.54	0.28	0.32
B3LYP	0.33	0.40	0.82	0.84
B3LYP-D3	1.13	1.32	0.83	0.95
CAM-B3LYP	0.37	0.49	0.97	0.94
ωB97X-D	1.20	1.44	0.61	0.80
M06-2X	0.76	0.91	0.79	0.91

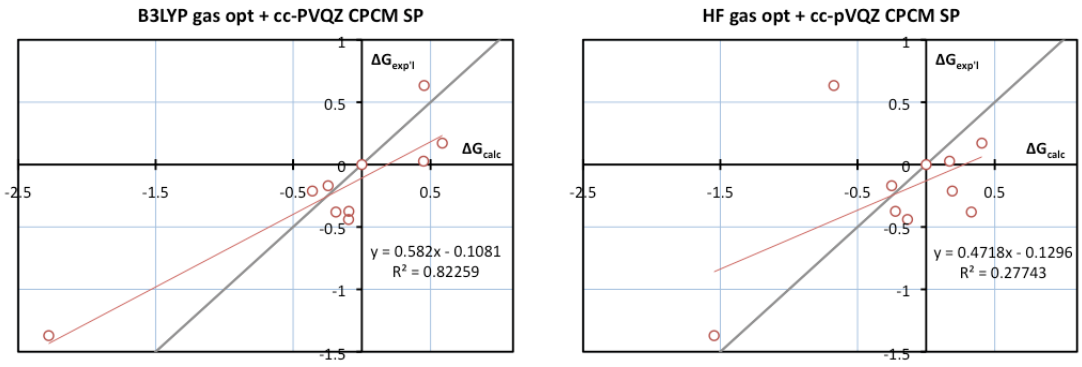


Figure 5-2: Plots of the experimental free energy differences for the molecular switches vs. the calculated differences. Energies are in kcal.mol^{-1}. The gray line is a plot of y = x, which is the slope that would ideally be calculated. The red line, whose equation is listed on the plot, is a linear fit to the data. These plots show that while the errors calculated using B3LYP and HF are low, there is relatively poor differentiation between the molecules, resulting in poor predictive power, especially for HF.

5.3.2. Solvent model

In order to investigate the effects of different implicit solvent models on the energies, single point calculations were repeated with the polarizable continuum model (PCM)[43] and the universal solvation model (SMD)[44] of chloroform at the B3LYP/cc-pVQZ level. While SMD solvation was consistently worse than CPCM solvation, PCM was very similar, with slightly improved PI and r^2 but slightly worse errors (Table 5-3). Based on these results, the CPCM solvent model was used for all future calculations.

Table 5-3: Statistical results from single point calculations using the listed solvent models. Each structure was optimized in the gas phase at the B3LYP/6-31+G(d,p) level before a single point calculation in chloroform with the cc-pVQZ basis set and the listed solvent model was performed.

Solvent Model	MUE	RMSD	r^2	PI
PCM	0.35	0.48	0.86	0.88
SMD	0.50	0.57	0.53	0.50
CPCM	0.33	0.40	0.82	0.84

5.3.3. Solution phase

To further investigate the role of solvation, the optimizations were repeated in CPCM chloroform, with electronic energy determined by a single point calculation, as before. This leads to an overall improvement in the linear correlations between theory and experiment, but in general does not reduce the actual errors (Table 5-4). This can be explained by many of the methods, in particular ωB97X-D, systematically overestimating the energy differences of the switches. This leads to an excellent rank-ordering of the compounds (a PI of 0.96), but relatively large MUE and RMSD as the linear correlation has a low slope of 0.23 relative to the ideal 1.0 (Figure 5-3). The performance of HF is most improved, with the lowest errors of any of the methods and a PI increasing from 0.32 for the gas phase calculations to 0.87 with the addition of solvent. This large improvement in predictive power results from only modest changes in energy for each molecule—on average

0.5 kcal.mol^{-1}—but given the small energy differences between the switches this is sufficient to drastically affect the results.

Table 5-4: List of statistical results from optimizations performed in CPCM chloroform. These solution phase calculations improve the linear correlation of most methods, but at a cost of higher absolute errors for some.

Functional	MUE	RMSD	r^2	PI
HF	0.25	0.36	0.85	0.87
B3LYP	0.42	0.56	0.88	0.98
B3LYP-D3	1.35	1.57	0.77	0.98
CAM-B3LYP	0.64	0.91	0.75	0.89
ωB97X-D	1.52	1.75	0.74	0.96
M06-2X	0.77	1.04	0.81	0.88

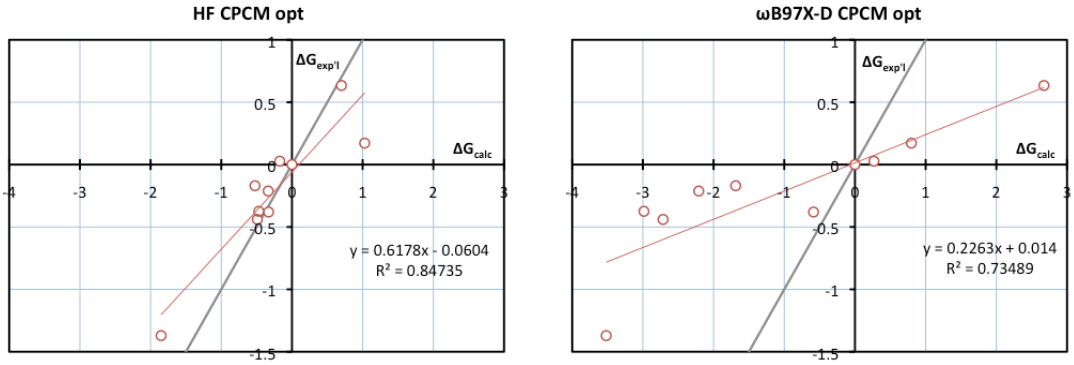


Figure 5-3: Plots of the experimental free energy differences for the molecular switches vs. the calculated solvent phase differences. Energies are in kcal.mol^{-1}. The gray line is a plot of y = x, which is the slope that would ideally be calculated. The red line, whose equation is listed on the plot, is a linear fit to the data. The inclusion of solvent to the optimization improves the correlations substantially, although errors remain high for ωB97X-D as the calculated values are much more spread out in energy than the experimental measurements.

5.4. Energies

5.4.1. Computational linear free energy relationships

Given the modest performance of the methods tested even with careful attention to basis sets and solvation, the accuracy of DFT methods in calculating the acidities of simpler systems was investigated. The ionization energies of substituted benzoic acids were thus considered in the manner of Hammett.[45] While extensive work has been directed towards accurate pKa calculations,[46] here the interest was in the performance of the DFT methods used for the full switches. The energies of the substituted benzoic acids and their

deprotonated forms were calculated at the B3LYP/cc-pVQZ//B3LYP/6-31+G(d,p) level in CPCM chloroform for both the optimization and SP. The $\Delta\Delta E_{Deprot}$ for the substituted benzoic acids (relative to benzoic acid) was plotted vs. the experimental ΔG of the switches, and an excellent linear correlation was found ($r^2 = 0.96$). Repeating these calculations for the deprotonation of aryl amides (Figure 5-4) resulted in a similarly strong linear trend ($r^2 = 0.95$). This indicates both that DFT calculations can correctly consider the acidities of these systems, and that the free energies of the switches can largely be correlated to the acidity of the amide NH.

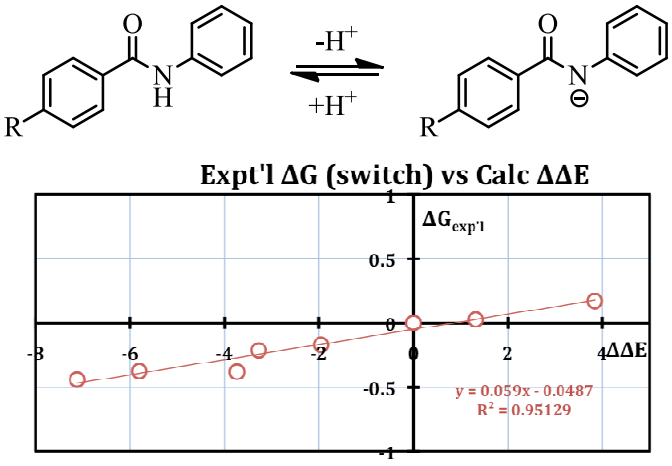


Figure 5-4: Top: Model system for calculating the relative deprotonation energies of substituted aryl amides. Bottom: Plot of the experimental free energy differences of the molecular switches vs. the calculated relative deprotonation energies of the amide system. There is an excellent linear correlation between these properties, implying that the DFT calculations are able to correctly predict acidities.

5.4.2. Single point energies of switches

As the DFT methods accurately model the acidities of the simplified systems, the full system was once again considered. While the previous analysis (Chapter 5.3) considered calculated free energies (which include zero point energy corrections and entropy), the electronic energy from the cc-pVQZ single point calculation can be analyzed separately. Comparing these energy differences to the measured free energy differences of the switches results in large errors, but excellent predictive power and correlation (Table 5-5). This is largely because the consideration of thermal effects serves to reduce the energy differences

between the switches, meaning the energies alone are along an erroneously large range (Figure 5-5).

Table 5-5: List of statistical results comparing the electronic energy differences calculated for the switches vs. the experimentally-measured free energies. These energy values are from single point calculations with the listed functional and the cc-pVQZ basis set in CPCM chloroform, using either the gas phase-optimized geometries or the solvent-optimized geometries. In all cases, these calculations result in excellent correlation and predictive index, but at the cost of relatively large errors.

	gas phase geometries				solution phase geometries			
Functional	**MUE**	**RMSD**	**r^2**	**PI**	**MUE**	**RMSD**	**r^2**	**PI**
HF	0.38	0.63	0.94	0.95	0.33	0.56	0.96	1.00
B3LYP	0.79	1.05	0.98	0.97	0.78	1.06	0.99	0.99
B3LYP-D3	1.81	2.20	0.85	1.00	1.71	2.06	0.88	1.00
CAM-B3LYP	0.98	1.25	0.98	0.99	0.95	1.25	0.98	1.00
ωB97X-D	1.54	1.97	0.76	0.96	2.03	2.43	0.78	0.96
M06-2X	0.76	1.11	0.96	0.96	0.73	1.09	0.96	0.93

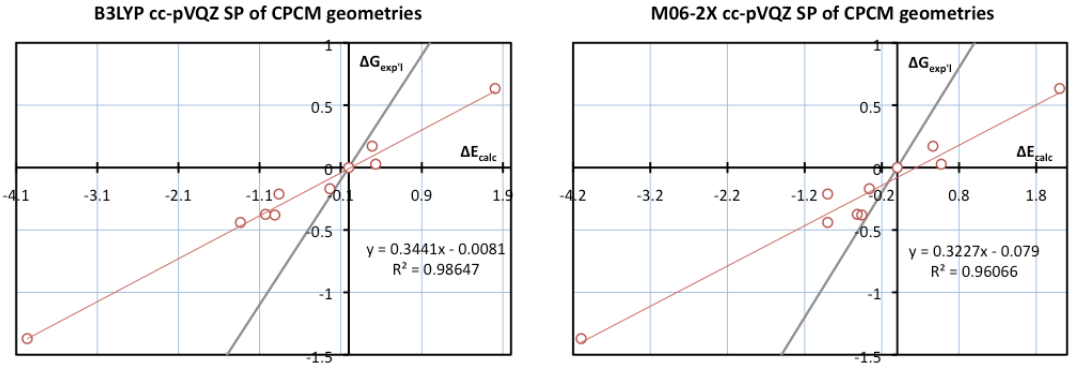


Figure 5-5: Plots of the experimental free energy differences for the molecular switches vs. the calculated energy differences with the cc-pVQZ basis set in CPCM chloroform. While ignoring thermal effects improves the correlation and PI of the calculations, it results in greater energy differences and larger errors.

5.5. Entropic effects

The energy calculations clearly indicate the importance of accurately considering entropic effects in calculating the free energy differences of the switches. Standard *ab initio* treatment of entropy relies on the simple harmonic oscillator (SHO) model of frequencies. For frequency ω, the vibrational entropy of such a system is:

$$S_{Vib} = R\left[\frac{h\omega}{k(e^{h\omega/kT}-1)} - \ln(1-e^{-h\omega/kT})\right]$$

where k is Boltzmann's constant, h is Planck's constant, T is the temperature, and R is the universal gas constant. As the frequency tends towards zero, the logarithmic portion of this equation diverges, meaning that small errors in calculating low vibrational modes can greatly influence the final calculated entropy.[47] Using the model of Grimme *et al.*,[34] these low modes are instead converted to the moment of inertia μ of a free-rotor:

$$\mu = \frac{h}{8\pi^2 \omega}$$

In the limit of small ω, the moment of inertia becomes large, so it is used in a reduced form:

$$\mu' = \frac{\mu B_{av}}{\mu + B_{av}}$$

where B_{av} is the average molecular moment of inertia, taken to be 10^{-44}kg.m^2. This leads to an overall rotational entropy for this rigid rotor harmonic oscillator (RRHO) of:

$$S_{Rot} = R\left[1/2 + \ln\left\{\left(\frac{8\pi^3 \mu' kT}{h^2}\right)^{1/2}\right\}\right]$$

In between the SHO treatment of large frequencies and the RRHO treatment of smaller ones, a damping function is applied which smoothes the transition. The choice of the cutoff frequency is somewhat arbitrary. The results described below are with a cutoff of 100 cm^{-1}, but using cutoff energies of 50 cm^{-1} or 150 cm^{-1} only changes the results by an average of < 0.05 kcal.mol^{-1} per switch.

This model was applied to the gas and solvent phase optimizations, resulting in a universal improvement in coefficient of determination and PI (Table 5-6). HF benefited the most from this entropic correction, with the gas phase r^2 improving from 0.28 to 0.82 with a corresponding increase in PI from 0.32 to 0.86. The largest correction was to the gas phase HF-calculated 2-pyridyl switch **9**, whose ΔG changed by 0.9 kcal.mol^{-1}. While both the "on" and "off" forms of that molecule were calculated to have 12 low vibrational frequencies, the lowest mode of the "on" form was 6.0 cm^{-1}, while for the "off" form it was

12.3 cm^{-1}. Thus the standard SHO model incorrectly calculates the "on" form to be more stable than in reality (due to an overly favourable TΔS term), an error which is corrected with the RRHO treatment.

Table 5-6: Statistical parameters calculated using the rigid rotor harmonic oscillator model for vibrational modes < 100 cm^{-1}. This modified treatment of entropy improves the coefficient of determination and PI for all of the functionals, but increases RMSD and MUE for some.

	gas phase optimization				solution phase optimization			
Functional	**MUE**	**RMSD**	**r^2**	**PI**	**MUE**	**RMSD**	**r^2**	**PI**
HF	0.24	0.31	0.82	0.86	0.24	0.33	0.95	0.97
B3LYP	0.42	0.55	0.98	0.97	0.43	0.49	0.90	0.98
B3LYP-D3	1.44	1.72	0.808	0.99	1.37	1.59	0.73	0.98
CAM-B3LYP	0.62	0.77	0.98	1.00	0.68	0.95	0.92	0.96
ωB97X-D	1.37	1.69	0.69	0.95	1.73	2.01	0.79	0.98
M06-2X	0.72	0.91	0.91	0.91	0.75	1.04	0.89	0.91

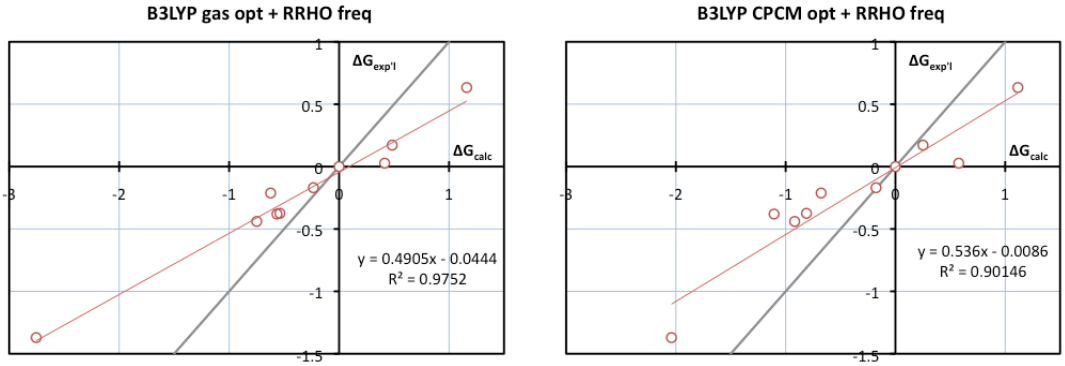


Figure 5-6: Plots of the experimental free energy differences for the molecular switches vs. the calculated differences when the RRHO model is used for low frequencies. This consideration of entropy improves coefficient of determination of all of the calculations.

For most other functionals and switches, the change in ΔG is quite small, generally < 0.2 $kcal.mol^{-1}$, and very rarely > 0.5 $kcal.mol^{-1}$. This is unsurprising as the vibrational differences between the "on" and "off" states of the system should be quite small, and thus errors generally cancel even without explicit consideration of the low modes. While the magnitude of the RRHO corrections to the solution-phase optimizations is similar to those for the gas phase (ca. 7 – 9 $kcal.mol^{-1}$ in each case), the corrections generally cancel for the solvent-optimized molecules, resulting in smaller net changes to ΔG.

5.6. Geometries

While solution-phase optimizations coupled with RRHO entropy and cc-pVQZ energies results in excellent rank ordering of the switches with all of the functionals tested, errors exceed the true energy differences of the system. As a result, the optimized geometries from the calculations were investigated more closely. Crystal structures have been reported for the *para*-nitro **5** and *para*-dimethylamino **6** switches.[22,24] Both show an approximately 10º dihedral angle for the central diphenyl moiety, with an approximately 50º out of plane twist of the phenyl ring adjacent to the H-bond caused by steric clashing with the methyl ester H-bond acceptor (Figure 5-7).

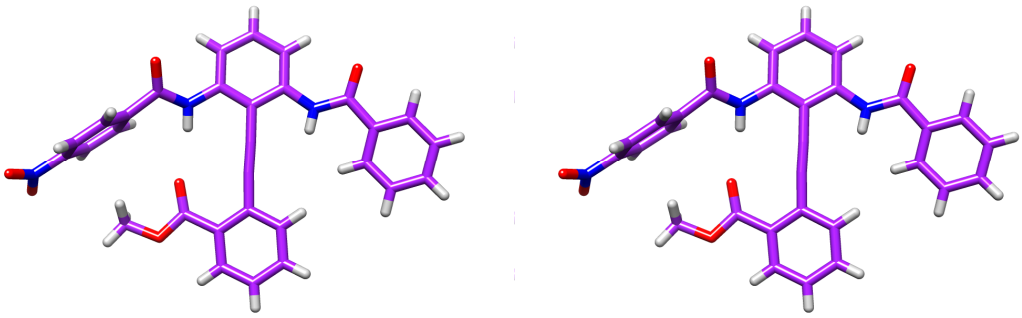


Figure 5-7: Stereo view of the crystal structure of the *para*-nitro molecule. The central DPA is nearly planar, while the phenyl ring bearing the nitro substituent is forced 50º out of plane due to steric clashing with the methyl ester.

To explore the calculated geometries, histograms of the DPA dihedral angles were plotted for each method. Significant disagreement within and among the functionals was found as to the preferred conformation (Figure 5-8). In the gas phase optimizations, each of the methods results in a multimodal distribution of geometries, with the plurality of optimizations resulting in angles of less than 30º but a significant minority at angles of 50º or greater. Optimizations in solvent gave less right-skewed histograms, but all methods still calculated some rotamers as roughly planar and others as significantly twisted out of plane. The most consistent results were with HF solvent optimizations, which calculated 40º - 50º dihedral angles for all but two molecules. While this is inconsistent with the solid-state data

and the other methods, these calculations were also the most accurate at reproducing the experimental free energies.

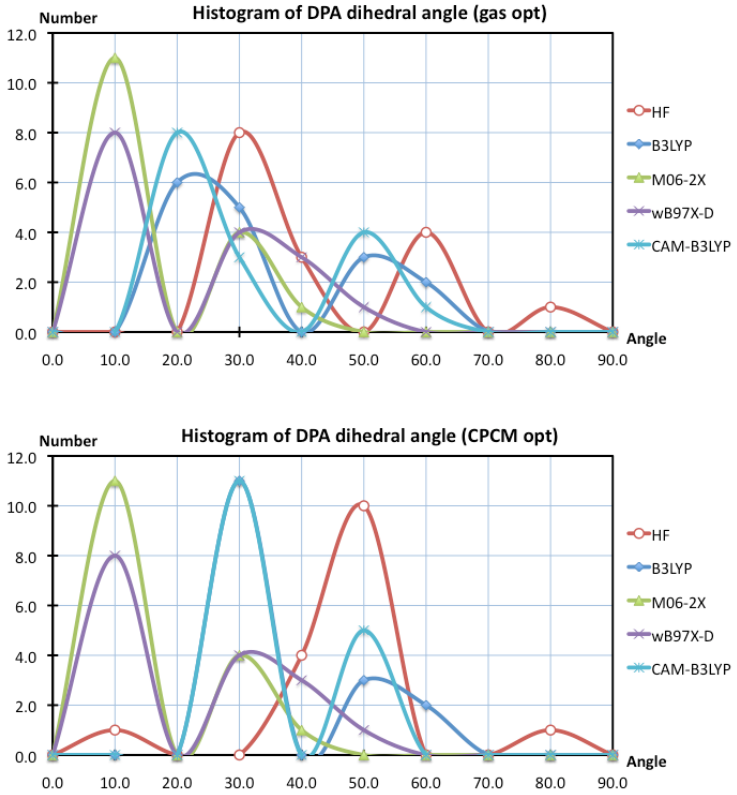


Figure 5-8: Histograms of the calculated DPA dihedral angle for each functional for optimizations performed in gas phase (top) or in CPCM chloroform (bottom). The addition of solvent reduces the skewing towards higher dihedral angles, but does not eliminate the multimodal profiles for most functionals. For the CPCM optimization, the highest peaks for B3LYP and CAM-B3LYP overlay at 30°.

Due to the relative flexibility of the system, several possible DPA dihedral angles can accommodate an NH•••O hydrogen bond by twisting the amide and ester groups relative to their phenyl rings. Overlays with the crystal structures reveal this effect. For instance, the solvent optimization of **6** with M06-2X results in a DPA dihedral of 24.5º compared to 11.5º in the crystal structure (Figure 5-9). As a result of this 15º shift, the benzamide rings of the calculated structure are both twisted approximately 30º relative to the DPA. This results in a CO•••N distance of 2.98 Å and an NHO angle of 153º. In the crystal structure, however, the phenyl ring adjacent to the amide is forced 50º out of plane by steric repulsion from the methyl ester, resulting in an H-bond distance of 3.10 Å and an angle of 145º.

Likewise for the *para*-nitro substituted **5**, ωB97X-D calculates a DPA dihedral of 0.3º. Instead of this resulting in one phenyl ring roatated significantly out of plane, however, the calculated structure shows equal 30º rotations of both benzamide rings and a 3.07 Å hydrogen bond. The crystal structure shows a hydrogen bond that is very similar in length (3.09 Å), but this is due to amide and ester carbonyls oriented in very different directions from those in the calculations (Figure 5-10).

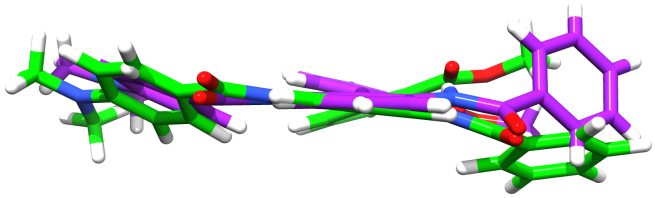


Figure 5-9: Overlay of the crystal structure of *para*-dimethylamino switch **6** (purple) with the structure calculated by M06-2X in CPCM chloroform (green). As seen from above, the crystal structure shows a more planar DPA than the calculation, which forces the phenyl ring significantly out of plane. The greater DPA twist in the calculation overcomes this strain, resulting in a different benzamide conformation.

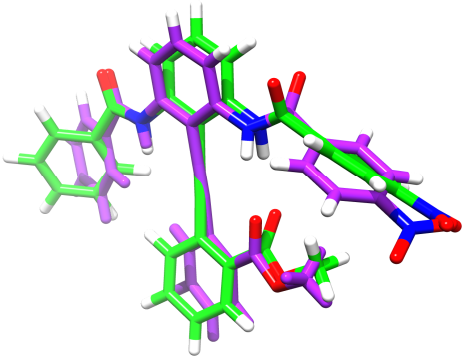


Figure 5-10: Overlay of the crystal structure of *para*-nitro switch **5** (purple) with the structure calculated with ωB97X-D in implicit CPCM chloroform. In the crystal structure, the methyl ester hydrogen bond acceptor is in the plane of the DPA, while the amide carbonyl rotates out of conjugation to position the NH for donating. In contrast, the calculation predicts both the amide and ester to rotate out of conjugation with the DPA, resulting in a nearly identical hydrogen bond length but a different conformation.

5.7. Constrained optimizations

5.7.1. DPA torsional profile

In an initial investigation, the accuracy of DFT methods for reproducing the torsional barrier of DPA was investigated (Figure 5-11). Fluorescence spectroscopy has measured the

barrier of rotation around the acetylene moiety to be 0.58 kcal.mol^{-1},[48] which has been successfully reproduced with MP2 calculations.[49] Previous DFT with B3LYP/6-31+G(d), however, calculates the barrier to be 0.81 kcal.mol^{-1}.[50] In this case, optimizations were performed where the DPA was constrained to be planar or perpendicular and the energy difference between these structures was taken to be the barrier height. Since the geometry of the full switches is fixed using the functional and the 6-31+G(d,p) basis set in CPCM chloroform, these methods were applied to the DPA system. All of the methods tested with the exception of B3LYP agree closely with the experimental barrier, with both M06-2X and cam-B3LYP reproducing the energy difference exactly (Table 5-7).

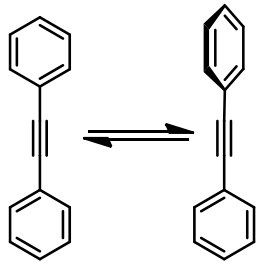


Figure 5-11: Rotation of the DPA moiety from the planar minimum energy to a 90º maximum energy.

Table 5-7: Calculated and experimental energy barriers for the 90º rotation of the phenyl rings of DPA using the listed functional and the 6-31+G(d,p) basis set. Consistent with literature investigations,[50] B3LYP overestimates the rotational barrier, but all other methods agree closely with the experimental values.

	Expt'l	**HF**	**B3LYP**	**cam-B3LYP**	**ωB97X-D**	**M06-2X**
ΔE (kcal.mol^{-1})	0.58	0.51	0.84	0.58	0.59	0.58

5.7.2. Full systems

The excellent results with the model DPA system led to consideration of the geometries of the switches. Optimizations were conducted using HF and B3LYP in which the central diphenyl was constrained to planarity, 30º, or 50º. These solution-phase optimized structures are not true minima, and so only their energies are compared to the experimental free energies of the switches (Table 5-8). For both functionals, the 50º dihedral angle results

in the lowest errors calculated so far, within 0.2 kcal.mol^{-1} for HF and 0.4 kcal.mol^{-1} for B3LYP. The coefficient of determination and PI for both methods are similarly excellent, with both methods perfectly ordering the compounds at 0º and B3LYP doing so at 50º as well.

Table 5-8: Statistical parameters calculated for optimizations in which the DPA is constrained to the listed dihedral angle. As the optimized structures are not true stationary points, the statistics are derived from the cc-pVQZ-derived energies for each molecule.

Functional	DPA dihedral	MUE	RMSD	r^2	PI
B3LYP	unconstrained	0.78	1.06	0.99	0.99
	0º	0.65	1.22	0.93	1.00
	30º	1.02	1.17	0.83	0.91
	50º	0.23	0.39	0.98	1.00
HF	unconstrained	0.33	0.56	0.96	1.00
	0º	0.48	0.87	0.91	1.00
	30º	0.49	0.81	0.90	0.97
	50º	0.20	0.22	0.92	0.92

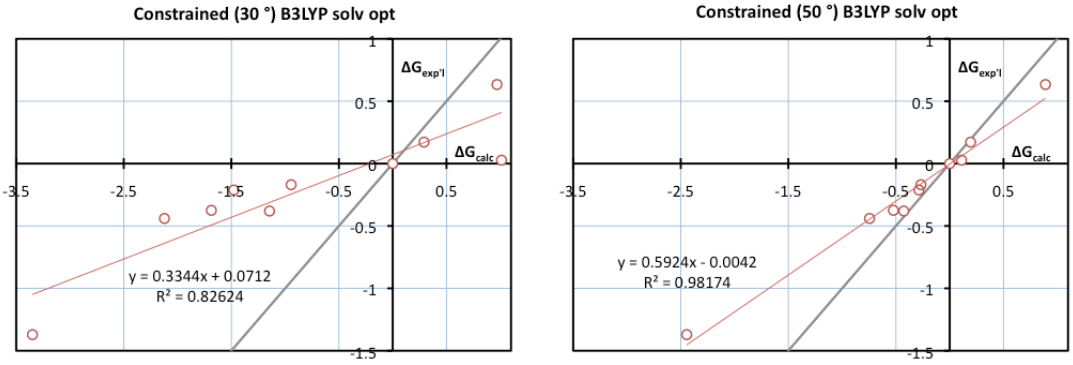


Figure 5-12: Plots of the experimental free energy differences for the molecular switches vs. the calculated energy differences at the B3LYP/cc-pVQZ//B3LYP/6-31+G(d,p) level when the DPA was constrained to the listed dihedral angle. With a 30º dihedral, the energy differences between the molecules are greater than at 50º, leading to greater errors and poorer linear correlation.

While the energy differences calculated at 0º and 50º dihedral angles were very close, when constrained to 30º the B3LYP calculations showed much greater spread in energies, with correspondingly larger errors (Figure 5-12). For the *para*-nitro molecule **5** for instance, comparison between the constrained optimizations, the unconstrained solution phase optimization, and experiment reveal some discrepancies (Table 5-9). All of the optimizations overestimate the H-bond distance, by up to 0.2 Å for the 50º optimization.

The conformational difference is likewise overestimated, but the consideration of entropic effects may reduce this somewhat (e.g. in the unconstrained case, the calculated energy difference is just -0.22 kcal.mol^{-1} with RRHO frequency analysis).

Table 5-9: Structural parameters for the *para*-nitro molecule **5** calculated by constrained optimizations and unconstrained optimizations with B3LYP in CPCM chloroform solvation.

	Unconstrained (26.5º)	**0º**	**30º**	**50º**	**Experiment**
ΔE (kcal.mol^{-1})	-1.03	-0.58	-1.69	-0.52	-0.37
N•••O (Å)	3.13	3.17	3.14	3.28	3.09
NCO angle (º)	158	160	157	152	161
Ar dihedral (º)	36.1	35.4	36.1	31.7	49.7

5.8. Discussion

The molecular switches of Hamilton *et al.* are an example of an organic system for which accurate quantum chemical calculations would be extremely useful. While synthesis of the molecules utilizes proven chemistry, each requires ca. eight steps to create, and two additional control molecules of similar complexity are needed to determine the free energy difference of the system. Calculations which could accurately predict the energies of novel switches could save considerable synthetic effort and allow efforts to be directed towards the most promising molecules. Once validated on this system, these computational methods could also be applied with some confidence to other hydrogen-bonded molecules.

These results indicate that "black box" DFT calculations, even with state-of-the-art functionals such as M06-2X and ωB97X-D, are extremely poor at modelling the conformational preferences of the switches. Thorough investigation of the calculations allowed some of the sources of error to be localized. One is spurious entropy components associated with low vibrational modes. In systems that are separated by small free energy differences, these entropic terms can lead to sizable relative errors. The RRHO correction to entropy greatly reduces these errors and adds essentially no computational cost to the calculations, so these results suggest that this correction should become standard. The

largest entropic corrections are applied to the very smallest frequencies, so the exact frequency at which the RRHO treatment is applied does not appear critical. Additionally, results were improved when optimizations were performed in an implicit solvent model. This adds some cost to the calculations, but in general it is small and is more than compensated for by the significant improvement in linear correlation and PI.

For the DPA-based switches, the most accurate results were ultimately achieved using the relatively simple functional B3LYP in combination with solution-phase optimizations, RRHO treatment of entropies, and large electronic single point calculations. Using these methods, RMSDs of < 0.50 kcal.mol^{-1} and a PI of 0.98 could be achieved, which are sufficient for guiding the development of related molecules. Every component of this new protocol is essential, and eliminating any resulted in poorer correlation with experimental data. To date, no “black box” protocol incorporates these critical factors, which calls into question calculations in the literature for which the claimed energy difference is only a few kcal.mol^{-1} and is calculated using uncorrected DFT methods. The performance of the more recent functionals ωB97X-D and M06-2X was comparatively worse than that of B3LYP. These methods were parameterized with larger datasets and incorporate more sophisticated treatment of non-bonded interactions, but nonetheless do not perform very well for this system.

These calculations, while still not perfect in modelling the switches, provide possible insight into the behaviour of these molecules. Principally, the torsional calculations and constrained geometry optimizations indicate that the system is quite flexible. In contrast to the crystal structures which show relatively planar structures, energies of the molecules were found to be roughly equivalent both when planar and when the DPA is twisted up to 50°. This is possible because in the calculations, the H-bonding amide and accepting carbonyl can themselves twist to accommodate an optimal H-bond length at a range of DPA

dihedrals. This could represent a failure of DFT, since these results are inconsistent with the X-ray data. Alternatively, the crystal structures may not reflect the true flatness of the potential energy surfaces. Instead, forces such as crystal packing effects or intermolecular hydrogen bonding may predominate in the solid state.[51]

The constrained geometry calculations raise further questions regarding the conformational equilibria. In the crystal structures, a pronounced twist in the phenyl ring attached to the donating amide is present due to steric clashing with the methyl ester acceptor. In the calculations, this is not present because the DPA twists to avoid this unfavourable interaction. Individual molecules twist to different extents, resulting in the multimodal histogram of DPA dihedral seen in the unconstrained cases. Restricting all of the molecules to a single angle generally results in poorer correlation with experiment, unless the constrained angle was 50°. It is unclear whether this is truly the optimal equilibrium conformation, or if this angle allows for fortuitous error cancellation, but it does provide the best correlation with experiment of any of the methods tested. These constrained calculations, however, do not allow for vibrations to be accurately calculated, and so this analysis does not include the entropic considerations that were found to be critical in the unconstrained cases. Nonetheless, they highlight how changes in the optimized geometry can greatly affect calculated energies. DFT-derived geometries are used for many composite methods such as CBS-QB3[52] and G3-B3[53] that use very large post-HF electronic energy calculations. While these methods have proven their accuracy for small systems, for larger molecules the DFT geometry may limit accuracy more than the electronic calculations themselves.

5.9. Conclusion and future work

Quantum chemical calculations were applied to a series of DPA molecular switches in which intramolecular hydrogen bonds template conformation. Poor correlation with

experiment was found with standard gas phase optimizations with a variety of DFT methods, but a new protocol incorporating solution-phase optimizations, large basis sets, and modified RRHO treatment of entropy gave much more accurate results. Using these methods, the switches could be ranked successfully in relative order of hydrogen bond strength, although errors remained between calculation and experiment for the absolute free energy differences of the molecules. Based on constrained optimizations, the DPA dihedral was not found to be essential to the energies of the system, in contrast to the clear preference seen in crystal structures. Together, these results benchmark the accuracy of DFT methods for medium-sized organic systems, and highlight the importance of the accurate treatment of solvation and entropy in quantum chemical calculations. The modified DFT procedure present here may be broadly applicable, and could result in greater accuracy in modelling systems for which non-covalent interactions are critical.

Additional benchmarking of the solvation and entropic corrections using large CCSD(T)/CBS databases could help confirm the utility of these methods. New functionals could then be developed which may improve on the poor performance of existing DFT methods. Future work should apply this modified DFT procedure to other systems of interest. For instance, the conformational preference of novel H-bonded peptidomimetics could be assessed *in silico* with greater accuracy than is possible with semi-empirical methods. This could allow for a more precise understanding of the conformational preferences of these molecules, which may aid future design efforts. Additionally, improved DFT modelling could aid the understanding of organocatalysis and allow greater confidence in the calculated structures of complexes. In any case, this study provides further evidence that “black box” modelling of structures cannot be considered truly reliable, especially in instances where the energy differences involved are slight.

5.10. Supporting information

Coordinates for all optimized structures and energies of all optimizations and single point calculations are attached as a CD to this thesis.

5.11. Supplementary description of computational methods

A brief outline of the computational methods used in this section is presented below:

5.11.1. Basis sets

6-31G: Pople split-valence basis set. Six contracted primitive Gaussians are used to describe the core orbitals. Valence orbitals are modeled in a double-ζ manner as the sum of three primitive Gaussians and an additional primitive Gaussian (of a different size).

6-31G(d,p): Polarized basis set augmenting the above with p orbitals on all H atoms and d orbitals on all non-H atoms. This increases the flexibility of the atom-centred atomic orbitals to describe molecular orbitals.

6-31+G(d,p): The above plus the addition of diffuse s and p orbitals on heavy atoms. These primitive Gaussians have small exponents and are correspondingly more spread out.

cc-pVTZ: Dunning's correlation consistent polarized basis set at the triple-ζ level. This basis set was optimized for calculations that include correlation.

cc-pVQZ: Like the above, but at the quadruple-ζ level. The Dunning basis sets are designed to scale smoothly to the CBS limit.

CBS: The complete basis set. It is not actually a basis set, but is rather an extrapolation from smaller calculations to the limit using an infinitely-large basis set.

5.11.2. Computational methods

Hartree-Fock: A method for approximating solutions of the Schrödinger equation using a single Slater determinant. Analytical solutions to the HF equations are unknown for multi-electron systems, so the energies are calculated in an iterative numerical manner.

Density Functional Theory: A method for solving the Schrödinger equation by considering functionals of the electron density of the system. The exact form of the electron density of molecules has not been determined, so different DFT methods use different approximations of the true value.

B3LYP: A hybrid DFT consisting of the three-parameter density functional of Becke with added Lee-Yang-Parr exchange. The total electronic exchange-correlation functional is a hybrid of the value from HF (20 %) with DFT terms (80 %).

CAM-B3LYP: Coulomb-attenuated modification to the B3LYP functional meant to address long-range interactions. In this functional, the fraction of HF exchange smoothly increases from 19 % at short range to 65 % at long range.

M06-2X: A hybrid DFT functional which incorporates 54 % HF exchange in addition to correlation values from several DFT approximations. The functional is parametrized to thermodynamic, kinetic, and noncovalent interactions from a reaction database.

ωB97X-D: A recently-developed functional by Head-Gordon which incorporates increasing amounts of HF exchange at long range. Additionally, an attractive term is included to describe dispersion interactions.

-D3: An empirical dispersion correction which is added onto DFT results.

5.12. References

(1) Sticke, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **1992**, *226*, 1143–1159.

(2) Meister, K.; Ebbinghaus, S.; Xu, Y.; Duman, J. G.; DeVries, A.; Gruebele, M.; Leitner, D. M.; Havenith, M. Long-range protein–water dynamics in hyperactive insect antifreeze proteins. *Proc. Natl. Acad. Sci.* **2013**, *110*, 1617–1622.

(3) Fan, E.; Van Arman, S. A.; Kincaid, S.; Hamilton, A. D. Molecular recognition: hydrogen-bonding receptors that function in highly competitive solvents. *J. Am. Chem. Soc.* **1993**, *115*, 369–370.

(4) Taylor, M. S.; Jacobsen, E. N. Asymmetric Catalysis by Chiral Hydrogen-Bond Donors. *Angew. Chem. Int. Ed.* **2006**, *45*, 1520–1543.

(5) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010**, *110*, 5023–5063.

(6) Burns, L. A.; Mayagoitia, Á. V.-; Sumpter, B. G.; Sherrill, C. D. Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *J. Chem. Phys.* **2011**, *134*, 084107–084132.
(7) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901–120901–12.
(8) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
(9) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
(10) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
(11) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.
(12) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. Parameterization of a B3LYP specific correction for non-covalent interactions and basis set superposition error on a gigantic dataset of CCSD(T) quality non-covalent interaction energies. *J. Chem. Theory Comput.* **2011**, *7*, 658–668.
(13) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Accounts* **2008**, *120*, 215–241.
(14) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(15) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(16) Vydrov, O. A.; Wu, Q.; Van Voorhis, T. Self-consistent implementation of a nonlocal van der Waals density functional with a Gaussian basis set. *J. Chem. Phys.* **2008**, *129*, 014106.
(17) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88–96.
(18) Ireta, J.; Neugebauer, J.; Scheffler, M. On the Accuracy of DFT for Describing Hydrogen Bonds: Dependence on the Bond Directionality. *J. Phys. Chem. A* **2004**, *108*, 5692–5698.
(19) Paton, R. S.; Goodman, J. M. Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *J. Chem. Inf. Model.* **2009**, *49*, 944–955.
(20) Riley, K. E.; Pitoňák, M.; Černý, J.; Hobza, P. On the Structure and Geometry of Biomolecular Binding Motifs (Hydrogen-Bonding, Stacking, X−H···π): WFT and DFT Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 66–80.

(21) Hujo, W.; Grimme, S. Comparison of the performance of dispersion-corrected density functional theory for weak hydrogen bonds. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13942.
(22) Jones, I. M.; Hamilton, A. D. Designed Molecular Switches: Controlling the Conformation of Benzamido-diphenylacetylenes. *Org. Lett.* **2010**, *12*, 3651–3653.
(23) Jones, I. M.; Hamilton, A. D. Anion-Dependent Switching: Dynamically Controlling the Conformation of Hydrogen-Bonded Diphenylacetylenes. *Angew. Chem. Int. Ed.* **2011**, *50*, 4597–4600.
(24) Jones, I. M.; Lingard, H.; Hamilton, A. D. pH-Dependent Conformational Switching in 2,6-Benzamidodiphenylacetylenes. *Angew. Chem. Int. Ed.* **2011**, *50*, 12569–12571.
(25) Dennington, R.; Keith, T.; Millam, J. *GaussView*; Semichem Inc.: Shawnee Mission KS, 2009.
(26) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. Gaussian 09, Revision B.01 **2009**.
(27) Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(28) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(29) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
(30) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
(31) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
(32) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.
(33) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *J. Comput. Chem.* **1983**, *4*, 294–301.
(34) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. – Eur. J.* **2012**, *18*, 9955–9964.
(35) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(36) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
(37) Wang, N. X.; Wilson, A. K. The behavior of density functionals with respect to basis set. I. The correlation consistent basis sets. *J. Chem. Phys.* **2004**, *121*, 7632–7646.
(38) Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.
(39) Prascher, B. P.; Wilson, A. K. The behaviour of density functionals with respect to basis set. V. Recontraction of correlation consistent basis sets. *Mol. Phys.* **2007**, *105*, 2899–2917.
(40) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
(41) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417–3423.
(42) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6*, 3850–3856.
(43) Mierts, S. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.
(44) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
(45) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
(46) Alongi, K. S.; Shields, G. C. Chapter 8 - Theoretical Calculations of Acid Dissociation Constants: A Review Article. In *Annual Reports in Computational Chemistry*; Ralph A. Wheeler, Ed.; Elsevier, 2010; Vol. Volume 6, pp. 113–138.
(47) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.
(48) Okuyama, K.; Hasegawa, T.; Ito, M.; Mikami, N. Electronic spectra of tolan in a supersonic free jet: large-amplitude torsional motion. *J. Phys. Chem.* **1984**, *88*, 1711–1716.
(49) Xu, D.; Cooksy, A. L. Ab initio study of the torsional motion in tolane. *J. Mol. Struct. THEOCHEM* **2007**, *815*, 119–125.
(50) Göller, A.; Klemm, E.; Egbe, D. M. Model systems for rod-like polyheteroarylethynylenes. *Int. J. Quantum Chem.* **2001**, *84*, 86–98.
(51) Dauber, P.; Hagler, A. T. Crystal packing, hydrogen bonding, and the effect of crystal forces on molecular conformation. *Accounts Chem. Res.* **1980**, *13*, 105–112.
(52) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
(53) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-3 theory using density functional geometries and zero-point energies. *J. Chem. Phys.* **1999**, *110*, 7650–7657.