

Copyright

by

Yan Bai

2009

**The Dissertation Committee for Yan Bai Certifies that this is the approved version
of the following dissertation:**

Progress in the Search for Ricin A Chain and Shiga Toxin Inhibitors

Committee:

Jon D Robertus, Supervisor

David Hoffman

Sean Kerwin

David Graham

Mona Mehdy

Progress in the Search for Ricin A Chain and Shiga Toxin Inhibitors

by

Yan Bai, B.Med.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2009

Dedication

This dissertation is dedicated to my loving parents, who are always supporting me, to my dear husband, Ben, and to our precious son, Nathan.

Acknowledgements

I would like to thank my advisor, Dr. Jon Robertus. He encouraged me to pursue my scientific interests, spent time to answer my naive questions, and, most importantly, inspired me by his ever-lasting enthusiasm for science.

Thanks to Art Monzingo and Kate Kavanagh, very patient and supportive crystallography mentors. Thanks for Suzaane and Beth, who put great efforts to assist this work. Thanks all other lab members, Matt, Karl, Divender, and especially Shuangluo for their great support and precious friendship.

Progress in the Search for Ricin A Chain and Shiga Toxin Inhibitors

Publication No. _____

Yan Bai, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Jon Robertus

Ricin and Shiga toxin type 1 are potent cytotoxins known as ribosome inhibition proteins, abbreviated RIPs. Proteins of this family shut down protein synthesis by removing a critical adenine in the conserved stem-loop structure of 28S rRNA. Due to its exquisite cytotoxicity, the plant toxin ricin has been used as a biological warfare agent. Although great achievement has been made on ricin research, including catalytic mechanism and structure analysis, there is still no specific treatment available for ricin exposure. In addition, ricin A chain inhibitors may also be useful against the homologous bacterial proteins shiga toxins, which are responsible for dysentery, and diseases related to food poisoning, including hemolytic uremic syndrome.

Previous study on RTA inhibitor search has provided a number of substrate analog inhibitors, all of which, however, are weaker inhibitors. Therefore, the goal of this work is to improve the binding affinity of known inhibitors and to discovery new scaffolds for inhibitor discovery and development. In this work, multiple approaches were employed for this purpose, including optimizing known inhibitors and searching new inhibitors by Virtual Drug Screening (VDS) and High Throughput Screening (HTS).

A number of new RTA inhibitors were discovered by these strategies, which provide a variety of pharmacophores for RTA inhibitor design, and also added a new line of evidence for VDS as an advanced technology for drug discovery and development.

Table of Contents

List of Tables	XI
List of Figures	XII
Chapter 1 Introduction	1
1.1 Significance of ricin study	1
1.2 Ricin biosynthesis	2
1.3 Biology of Shiga toxin	4
1.4 Structure of ricin and shiga toxin	7
1.5 Cell entry and transport of ricin and shiga toxin.....	12
1.6 Catalytic mechanism	16
1.7 Ricin inhibitor identification	22
1.8 Strategies for RTA inhibitor design and discovery	25
Chapter 2 Continuing structural studies on purine and pterin based RTA inhibitors	30
2.1 Introduction.....	30
2.2 Materials and methods	30
2.2.1 RTA expression and purification	30
2.2.2 Determination of RTA and RTA inhibitor activity.....	31
2.2.3 Crystallization of RTA.....	35
2.2.4 Inhibitor soaking and co-crystallization.....	35
2.2.5 X-ray data collection and structure determination.....	36
2.3 Results and discussion	37
Chapter 3 RTA inhibitors identified by Virtual Drug Screening (VDS)	
3.1 Introduction.....	49
3.2 Materials and methods	52
3.2.1 Compound database	52
3.2.2. Receptor file preparation.....	52
3.2.3 Computer workstation.....	53
3.2.4 Docking with ICM	53

Vita.....	147
-----------	-----

List of Tables

Table 1:	Seven more RTA inhibitors derived from purine and pterin families.....	40
Table 2:	SAR analysis for purine/pterin-based RTA inhibitors.....	48
Table 3:	Evaluation of GOLD and ICM ranking performance using known RTA inhibitors.....	67
Table 4:	Top 10 hits from Sigma-Aldrich library predicted by ICM.....	70
Table 5:	PBA data collection and refinement statistics	75
Table 6:	PBA derivatives	80
Table 7:	ICM docking results for amino acid coupled PBA.....	89
Table 8:	Hits identified from Chembridge library by VDS	91
Table 9:	Screened libraries by cell-based HTS at the NSRB.....	113
Table 10:	Hits selected from the cell-based HTS.....	114

List of Figures

Figure 1:	Schematic diagrams of RIPs	6
Figure 2:	Ricin X-ray structure	9
Figure 3:	X-ray structure of shiga toxin	10
Figure 4:	Superposition of the X-ray structures of RTA and Stx A chain	11
Figure 5:	Ricin and shiga trafficking in mammalian cells	15
Figure 6:	A diagram of the SRD and its X-ray structure	19
Figure 7:	RTA active site upon FMP binding.....	20
Figure 8:	Proposed mechanism for RTA depurination of rRNA	21
Figure 9:	Binding modes for representatives of adenine, guanine and ptein based RTA inhibitors	24
Figure 10:	Illustration of the design of portential strong RTA inhibitors	29
Figure 11:	Cartoon of the luciferase assay	33
Figure 12:	Conversion of 7P9DG in the X-ray structure	41
Figure 13:	The active site of RTA•7M9DG crystal structure	42
Figure 14:	Superposition of RTA•9DG, RTA•7M9DG and RTA•7HP9DG	43
Figure 15:	X-ray structure of RTA•7CP	45
Figure 16:	The correlation plot of eHiTs scores and IC ₅₀ for known RTA inhibitors	60
Figure 17:	The correlation plot between GOLD score and ICM score for Chembridge fragment library	65
Figure 18:	The correlation plot between GOLD score and molecular weight of the control ligands used in the test.....	66
Figure 19:	Inhibition of RTA by PBA.....	73

Figure 20:	Stereo view of the omit electron density map for PBA in RTA•PBA	74
Figure 21:	The Ligplot of interaction between PBA and RTA	76
Figure 22:	Superposition of the ICM predicted PBA binding pose with the X-ray structure of RTA•PBA	77
Figure 23:	Superposition of the active sites of crystal structure for RTA•PBA and RTA•PTA.....	78
Figure 24:	The correlation between ICM score and IC ₅₀ for PBA derivatives..	82
Figure 25:	The correlation of GOLD score and ICM score	83
Figure 26:	Consistent docking poses for pyrimidine hits by GOLD and ICM..	84
Figure 27:	The model of histidine coupled PBA bound at two pockets.....	86
Figure 28:	Protein synthesis altered by compound 5225540.....	94
Figure 29:	The stereo view of omit difference density for the RTA•5225540..	98
Figure 30:	The stereo view of the superposition of RTA•7543758 and apo RTA	99
Figure 31:	Stereo view of the omitted difference map of RTA•7780474	100
Figure 32:	Cartoon of cell-free HTS assay	105
Figure 33:	Cartoon of cell-based HTS assay.....	112
Figure 34:	Compound 2 dose response curve in cell-based assay.....	117
Figure 35:	384-plate layout for NCC library HTS	121
Figure 36:	The honokiol structure and its activity with or without BSA.....	122
Figure 37:	CD experiment data for RTA and RTA with different compounds	125
Figure 38:	RTA T _m changes by compounds in DSF	128
Figure 39:	Representation ITC data for the interaction of 7CP and RTA	133

Chapter 1: INTRODUCTION

1.1 Significance of ricin study

Ricin is a potent plant toxin produced in castor oil plants, *Ricinus communis*, and it is naturally used as a weapon by those plants to defend themselves from foreign attacks. Ricin is mainly synthesized and stored in the castor seeds, where it constitutes up to 5% of the total protein content of the seeds (Marsden, Smith et al. 2005). During the industrial production of castor oil, the oil is extracted after pressing the seeds, leaving a waste mash of which ricin is a significant constituent and can be extracted through a simple and inexpensive process. Ricin has a half lethal dose, an LD₅₀, of ~1µg/kg body weight for mice, rats, and dogs and 0.1µg/kg for rabbits. For humans, the LD₅₀ of ricin could be as low as 3 µg/kg body weight if injected (Gonatas, Gonatas et al. 1998).

Because of its toxicity and availability, ricin has been used as a lethal biochemical agent. The most famous ricin attack was in the 1978 “umbrella-tip” murder of Georgy Markov, who was a journalist in London. He died three days later, which was caused by less than 500 µg ricin according to the estimation (Ellenhorn 1988). After this, several more ricin attacks have been widely reported. More recently, ricin was detected in the mail at the White House in Washington, D.C. and in an office mailroom of U.S. Senate Majority Leader Bill Frist’s office in 2003 and 2004 respectively. Although ricin is not an infectious bio-terror hazard such as anthrax, it still remains a public threat because it is a deadly poison that can be easily obtained. However, there is still no effective treatment available for ricin exposure, which has led to a worldwide ricin inhibitor search.

Years of research on ricin biology and biochemistry have provided a scientific basis for ricin inhibitor design. Ricin is a member of type 2 ribosome inactivating proteins, abbreviated RIPs (Olsnes and Kozlov 2001). RIPs are enzymes that inhibit protein synthesis by removing one key adenine residue from rRNA (Stirpe and Battelli 2006). In nature, RIPs are widespread in plants, and are present also in fungi, bacteria, and at least one alga (Girbes, Ferreras et al. 2004; Prestle, Schonfelder et al. 1992; Ishizaki, Megumi et al. 2002; Barbieri, Polito et al. 2006). There are two major groups of RIPs, type 1 and type 2, depending on the chain composition. Type 1 RIPs have a single chain, strongly basic proteins of approximately 30 kDa with enzymatic activity; type 2 RIPs consist of two chain types, one of approximately 30 kDa with enzymatic activity and one with lectin properties (Fig. 1). Plant RIPs have a single lectin chain of around 30 kDa, but bacterial toxins typically have a pentamer of smaller lectins. The plant lectins are usually specific for sugars with the terminal free D-galactose structure, thus facilitating entry of the RIPs into cells (Barbieri, Battelli et al. 1993). Type 2 RIPs are the focus of greatest interest because of their ability to kill cells, whereas type 1 RIPs inhibit only cell-free protein synthesis but are relatively non-toxic to cells and animals. Ricin and shiga toxin (Donohue-Rolfe, Acheson et al. 1991), as representatives of highly toxic plant and bacterial type 2 RIPs, have been extensively studied.

1.2 Ricin biosynthesis

Ricin is synthesized in the oil-storing endosperm cells of maturing *Ricinus communis* seeds, where it accumulates in protein storage vacuoles (Lord, Roberts et al. 1994). Ricin A chain and B chain are originally synthesized together as a single precursor protein, which is processed into two chains

during biological development (Butterworth and Lord 1983) (Fig 1b.). The gene of the ricin precursor encodes a preproprotein of 576 amino acid residues, consisting of a 35-residue N-terminal extension, the first 26 residues of which are a signal sequence (Ferrini, Martin et al. 1995), followed by the 267 residues of RTA, a 12-residue linker peptide, and the 262 residues of RTB (Lamb, Roberts et al. 1985). During biosynthesis, the ricin precursor reaches and enters the lumen of the ER under the guidance of the N-terminal signal sequence, which is, in turn, immediately removed by a signal peptidase, leaving the proricin portion segregated in this cellular compartment (Roberts and Lord 1981). The proricin is then core-glycosylated and oxidized, resulting in the formation of five disulfide bonds within the protein (Lord 1985). Four of the disulfide bonds are formed within the RTB and one forms between the RTA and RTB as a linker. Then, the core-glycosylated, disulfide cross-linked proricin moves through vesicular transport from ER, via the Golgi stack, to the vacuole (Lord 1985; Lord 1985); the movement is accompanied by several largely uncharacterized posttranslational modifications (Harley 1985; Lord 1985; Frigerio, Jolliffe et al. 2001). The 12-residue peptide linking the RTA and RTB sequences carries a sequence-specific determinant for vacuolar targeting (Frigerio, Jolliffe et al. 2001). When proricin reaches the vacuole, the nine residues of the 35-residue N-terminal leader sequence and the 12-residue linker peptide between RTA and RTB are removed by acidic endoproteases. This process generates, finally, the disulfide-linked heterodimeric native toxin (Harley 1985), which is stored inside the vacuole and ready to play its biological role of attacking cells and killing ribosomes.

Ricinus communis ribosomes are sensitive to the catalytic action of ricin A chain, but the ricin biosynthesis pathway evolved to ensure the enzyme and

substrate never encounter one another, therefore keeping its own protein synthesis safe. In addition, the RTA component of proricin is not catalytically active until it is released from RTB by disulfide bond cleavage (Richardson, Westby et al. 1989). This ensures that any proricin misdirected into the cytosol would not inevitably kill its own ribosomes. Furthermore, biologically active heterodimeric ricin is stored within vacuoles in the mature *R. communis* seeds, an intracellular compartment from which it never escapes (Tully and Beevers 1976; Youle and Huang 1976). Ricin and other proteins stored in the seeds will be rapidly and completely degraded by vacuolar proteases during the germination to provide a source of amino acids for protein synthesis. However, before germination, if the seeds are attacked by foreign intruders, ricin will be released from broken vacuoles to defend the plants from devastation, which is a part of plant bio-defense mechanism.

1.3 Biology of shiga toxin

Shiga toxin (Shiga 1898) is another type 2 RIP, present in certain bacteria. The bacterial strains that produce shiga toxin are generally pathogenic, and are responsible for a number of diseases such as hemorrhagic colitis, neonatal and adult diarrhea, and hemolytic uremic syndrome.

The toxin, produced by *Shigella dysenteriae*, is abbreviated as Stx. The shiga toxins produced by *E. coli* are named shiga-like toxins, Slt, and are categorized into two groups, Slt-1 (shiga like toxin 1) and Slt-2 (shiga like toxin 2). The amino acid sequences of these toxins are highly conserved: For the A subunits, Slt-1 differs from Stx by only 1 amino acid. Slt-2 shares 56% sequence similarity with Slt-1 (Calderwood, Auclair et al. 1987). All members of shiga toxin

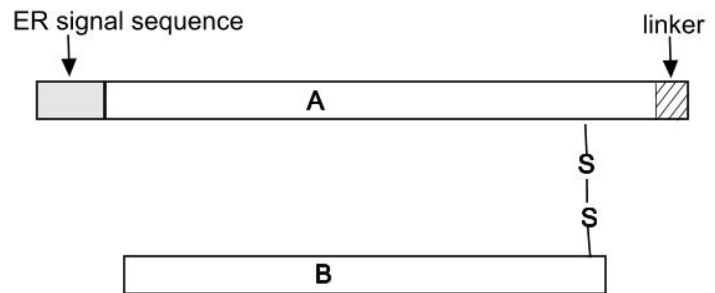
family are bipartite molecules composed of a single enzymatic A subunit and a multimeric receptor-binding B subunits (Olsnes, Reisbig et al. 1981; Donohue-Rolfe, Keusch et al. 1984). The mature A subunit of shiga toxin, StxA, is composed of 293 amino acids with a molecular weight about 32 KDa (Strockbine, Jackson et al. 1988), and the mature B subunit consists of 69 amino acids about 7.7 KDa (Seidah, Donohue-Rolfe et al. 1986). Like ricin, the A polypeptides of the shiga toxin family are activated by proteolytic processing. A bacterial protease nicks the shiga toxin and Slt A subunits into an enzymatic A1 fragment (253 amino acids), and a carboxyl terminal A2 fragment (40 amino acids). The A2 segment remains linked with A1 by a single disulfide bond between C242 and C261 in A1 and A2 respectively until the enzymatic fragment is released in ER and enters the cytosol of mammalian cell. (Olsnes, Reisbig et al. 1981; Pelham 1990; Sandvig, Ryd et al. 1994) The five small B polypeptides are noncovalently associated with the A2 portion of the A subunit in the holotoxin (Fig.1C). Previous studies have shown that shiga toxin B subunit is a cell surface binding moiety, and the principle functional receptor of shiga toxin is a membrane glycolipid, globotriaosylceramide (Gb3) (Keusch and Jacewicz 1977; Karmali, Petric et al. 1985).

Although the entire sequence similarity is low for ricin and shiga toxin, there are two regions of limited sequence homology between them within the A chains. These conserved regions form the toxin active site that endows shiga toxins with the same enzymatic function as ricin. Therefore, inhibitors against ricin A chain are also potential antidotes for shiga toxin.

A. Type 1 RIPs



B. Type 2 RIPs -- Ricin



C. Type 2 RIPs -- Shiga Toxin

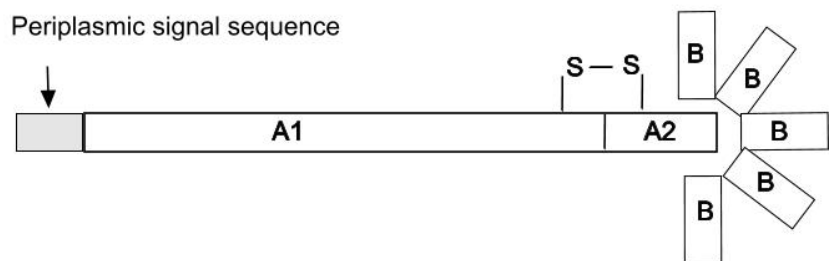


Figure 1. Schematic diagrams of RIPs. A is the enzymatic chain, B is cell-surface binding components. Signal sequence is in grey. -S-S- illustrates disulfide bond.

1.4 Structure of ricin and shiga toxin

The X-ray structure of ricin was the first to be determined in the RIP family (Montfort, Villafranca et al. 1987). The holoenzyme is a globular, glycosylated heterodimer, containing the A chain (RTA) and the B chain (RTB). These two chains are joined together by a disulfide bond (Fig. 2). Secondary structure of RTA includes a six-stranded β -sheet, and a two-stranded β -sheet. The protein can be divided into three folding zones, or domains. . The first domain is a compact module formed by six β -strands and first two α -helices. The central domain is made up of five helices, including the most part of active site. This part contains the key active site residues Glu177 and Arg180, which lie on the turns of a helix and expose to the solvent of the active site. The third domain consists of a two-stranded antiparallel β -sheet and a helix, which includes another part of the active site and is responsible for interacting with RTB (Robertus 1991; Rutenber, Katzin et al. 1991). These secondary structure elements in the N-glycosidase domain are highly conserved for all RIPs, although the primary sequences show only limited sequences identity. The active site residues are invariant, including Tyr80, Tyr123, Glu177, Arg180, and Trp211 (Ready, Katzin et al. 1988; Monzingo, Collins et al. 1993).

The ricin B chain has a bilobal structure in which the two domains are homologous and arose by gene duplication (Rutenber and Robertus 1991). Each domain contains three structurally conserved subdomains, named α , β , γ . Each subdomain consists of about 40 amino acid residues. It is believed that these subdomains existed as freely associating trimers prior to their fusion to form the modern galactose binding domains. In RTB, only the 1α and 2γ subdomains

retain lectin activity. The sugar binding pockets are formed by a sharp bend in the backbone together with an aromatic residue, usually Trp, which provides the binding platform for the hydrophobic face of the bound sugar (Rutenber and Robertus, 1991).

For shiga toxin B chains, crystal structures show that each B subunit has one α -helix and six β -strands (Fig. 3). These five B subunits form a pentameric ring with the helices inside and β sheets outside. The carboxyl-terminal tail of the A subunit lies in the center of the ring, and the rest of the A-subunit lies on one side of the B-pentamer (Fraser, Cherniaia et al. 1994; Fraser, Fujinaga et al. 2004). As mentioned above, the A-subunit of the Stxs can be proteolytically cleaved, but the two portions, A1 and A2 remains covalently associated through a disulfide bond until it reaches ER. This disulfide bond (between Cys242 and C261 in pdb 1DM0) is located near the residues that form the N-glycosidase active site, which places M260 right outside the active site gate to block it. In this conformation, the holotoxin can stay enzymatically inactive (Fraser, Cherniaia et al. 1994; Fraser, Fujinaga et al. 2004).

The folding of Stxs A subunits resemble the structure of RTA: the N-terminal portion of Stxs starts with six-stranded β -sheets, followed the seven helices and two more β sheets in the A1 portion (Fig. 4). A2 is mainly composed of a short helix, which is the part that lacks on RTA. The enzymatic active site of Stxs is essentially identical to the site of RTA, with the exception of the side chain of Met 260 (1DM0.pdb) from A2 which inserts in the active site, as another strategy to keep the holoenzyme inactive (Reisbig, Olsnes et al. 1981; Fraser, Cherniaia et al. 1994).

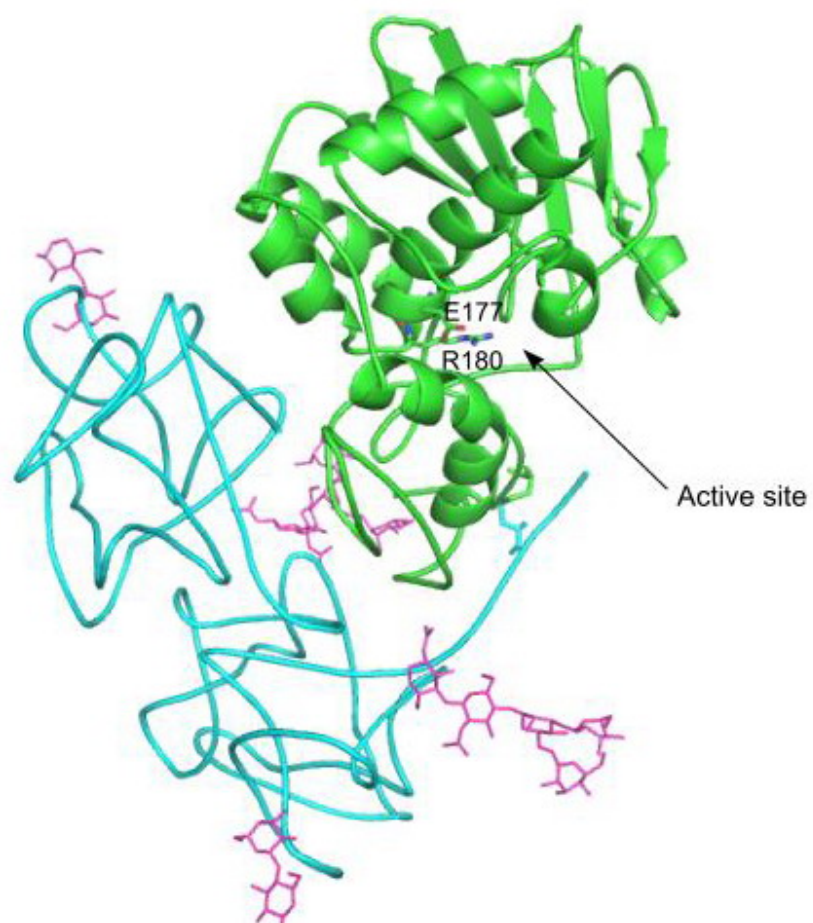


Figure 2. Ricin X-ray structure (2AAI.pdb): RTA in green, RTB in cyan, bound sugar molecules in purple, active site residues labeled as E177 and R180, enzymatic active site indicated by the arrow.

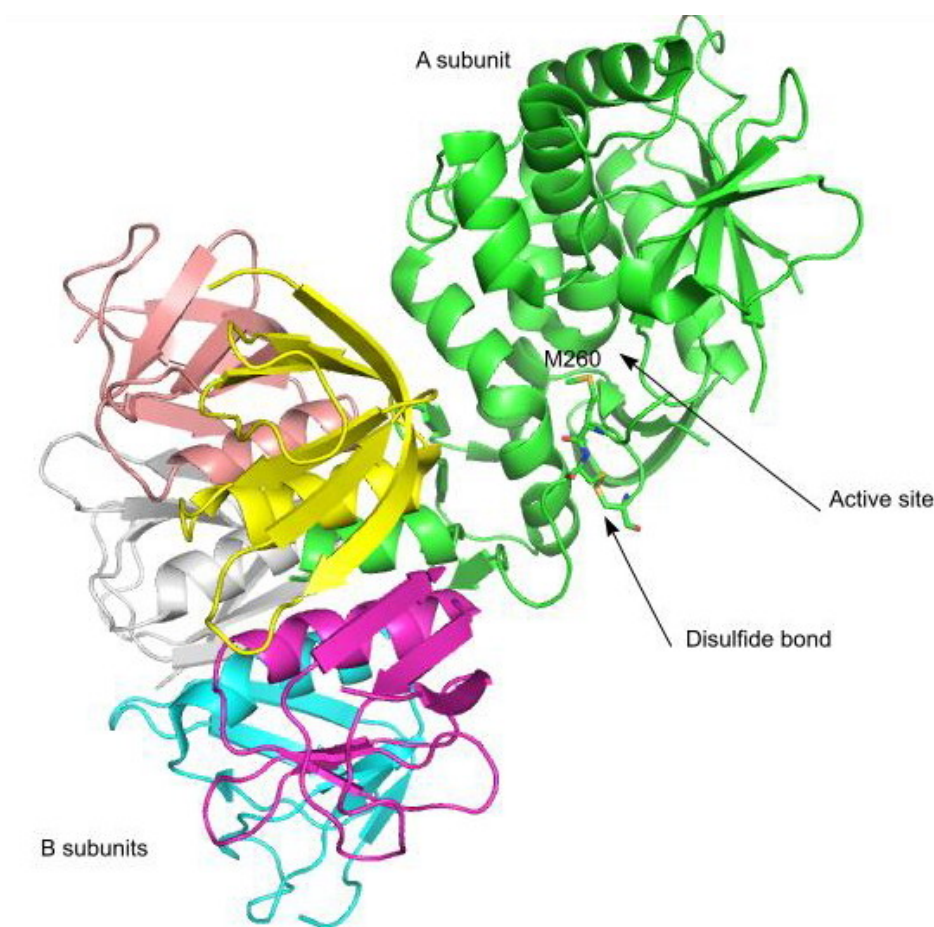


Figure 3. X-ray structure of shiga toxin (1DM0.PDB). The enzymatic active chain (A) is in green, five B chains are in pink, yellow, purple, cyan, and grey respectively. Active site and disulfide bond are shown with arrows. The residue M260 that blocks the active site is labeled.

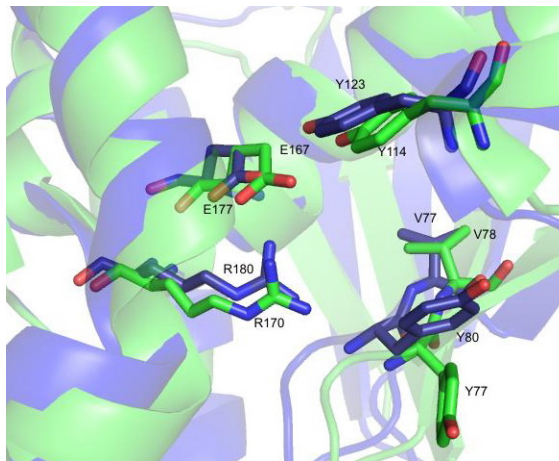
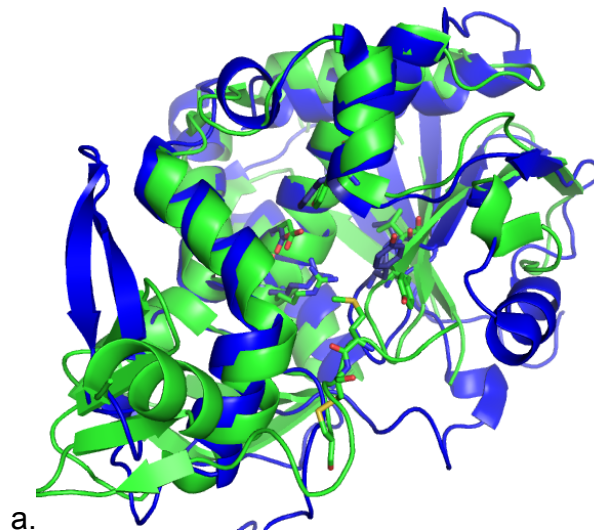


Figure 4 (a). Superposition of the X-ray structures of RTA (blue) and Stx (green) Despite a low sequence similarity between RTA and Stx A chain, they fold in the same way. (b). Superposition of the active site residues of RTA (blue) and Stx (green). The active site residues of RTA and Stx A chain are identical, shown as labeled.

1.5 Cell entry and transport of ricin and shiga toxin

Like other catalytically active protein toxins, ricin needs to bind the target cell surface by its B chain in order to enter that cell. The substrates of ricin B chain are β -1,4-linked galactose residues, which are widely present on mammalian cell surface of glycoproteins and glycolipids. Therefore ricin can readily bind to and enter most cell types (Sandvig, Olsnes et al. 1979).

Shiga toxins enter the cell upon B chain binding to neutral glycolipid globotriasylceramide such as Gb3 (Sandvig, Olsnes et al. 1979). Stx2e, a variant of Stx2, binds preferentially to the globotetraosylceramide Gb4. (Fraser, Fujinaga et al. 2004). These different binding properties lead to the ability of the toxin to selectively target different cells.

Both ricin and shiga toxins enter cells by endocytosis through multiple pathways, including clathrin-dependent endocytosis (van Deurs, Pedersen et al. 1985) and clathrin-independent mechanisms (Hinshaw and Schmid 1995; Llorente, Rapak et al. 1998; Simpson, Smith et al. 1998; Sandvig, Grimmer et al. 2002). After entering cells, ricin is initially delivered to early endosomes, from where the majority of ricin is either recycled back to the cell surface or sent to lysosomes for subsequent proteolytic degradation (Sandvig and Olsnes 1980). However, a small portion of ricin molecules are able to go through the Golgi complex by a pathway that is independent of late endosomes, as has been confirmed by several lines of biochemical evidence (van Deurs, Tonnessen et al. 1986; Lombardi, Soldati et al. 1993; Riederer, Soldati et al. 1994; Simpson, Dascher et al. 1995; Sandvig and van Deurs 1996; Iversen, Skretting et al. 2001). Recent reports suggest that specific proteins such as Rab6' and cognate SNAREs are involved in the direct early endosome-to-TGN transport step for

shiga toxin trafficking, and it is possible that ricin may also exploit the direct route from early endosomes to the Golgi.

In order to exert a toxin effect, ricin needs to traffic to the ER. This is the only route from which it can reach its site of action, the cytosol, because the ER is the only organelle that contains pre-existing protein-conducting channels through which the toxin can exit into the cytosol. Both ricin and shiga toxin can be transported into the ER in a retrograde fashion (Sandvig and van Deurs 1996; Rapak, Falnes et al. 1997). It is known that some AB toxins, like CT, LT, and ExoA, have a C-terminal ER-retrieval signal and are apparently returned to the ER via the KDEL receptor (Chaudhary, Jinno et al. 1990; Pelham 1990; Nilsson and Warren 1994; Cieplak, Messer et al. 1995; Miesenbock and Rothman 1995). Ricin might use this method to return the ER although it doesn't have its own KDEL sequence. It was suggested that RTB's galactose binding ability may be used to bind galactose-modified, KDEL-tagged proteins traveling from the TGN to the ER; this has been supported by the observation that disruption of RTB's intracellular galactose binding ability eliminates ricin's toxic effects (Pelham, Roberts et al. 1992). In the case of shiga toxin, there is evidence that a COPI-independent Rab6-dependent retrograde transport route was used (Girod, Storrie et al. 1999; White, Johannes et al. 1999).

Once ricin gets into the ER, RTA dissociates from RTB through the breakage of the disulfide bond by protein disulfide bond isomerase (Majoul, Ferrari et al. 1997). and becomes partially unfolded. This partial unfolding enables RTA to be expelled from the ER into cytosol by the quality control system that operates there (Ellgaard, Molinari et al. 1999). This system is known as ER-associated protein degradation (ERAD); it can detect newly synthesized

proteins that are not properly folded in their biologically active conformation, and send them into cytosol for proteolytic degradation. The notion that the Sec61p protein complex plays a role in the retrograde toxin transport from the ER to the cytosol, is supported by the observation that RTA was co-immunoprecipitated with Sec61 α (Simpson, Roberts et al. 1999; Wesche, Rapak et al. 1999).

Once transported to the cytoplasm by the ERAD system, degradation is carried out by proteasomes, which usually use the ubiquitylation pathway. Clearly, a proportion of RTA must be able to escape the cytosolic degradation system and refold into its native conformation. One possible explanation for escaping degradation might be that RTA has a very low lysine content, which is the tag for ubiquitylation process, and it has been proposed as an evolutionary strategy to allow the protein to avoid ERAD (Hazes and Read 1997). However, little is known about the way that partially unfolded RTA is capable of refolding into its active conformation. There is some suggestion that RTA might use the ribosome, its substrate, as a refolding chaperone (Argent, Parrott et al. 2000). The trafficking of ricin and Shiga toxin is illustrated in Fig. 5.

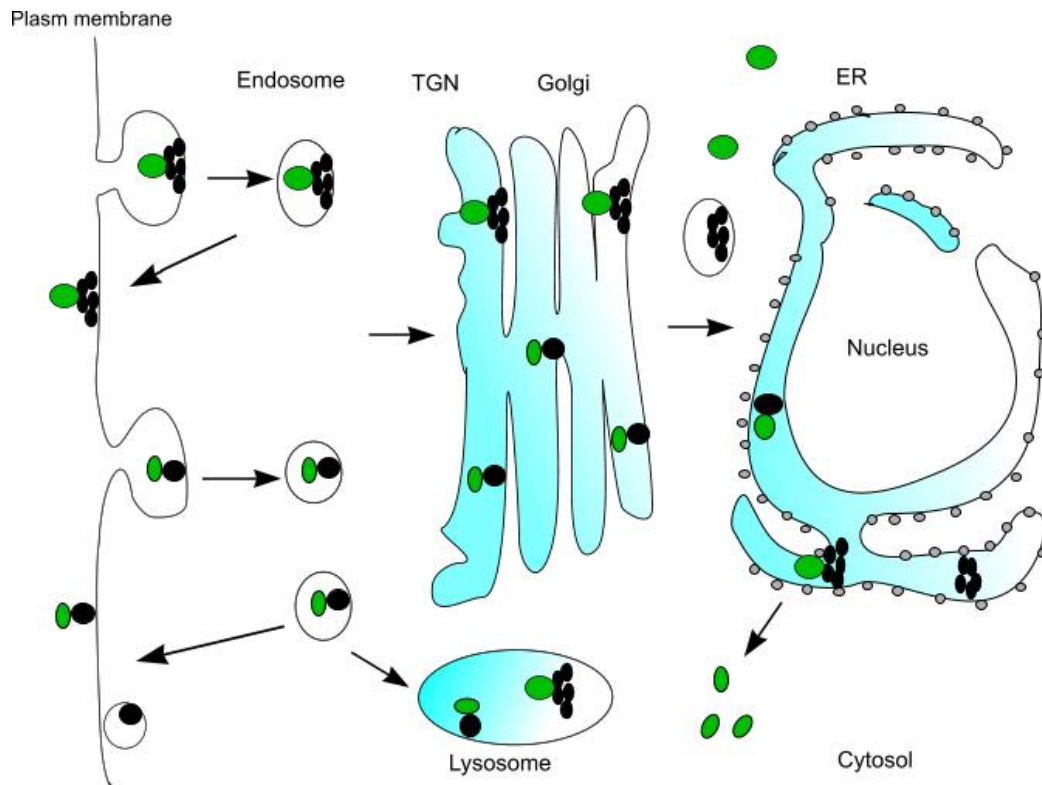


Figure 5. Ricin and shiga trafficking in mammalian cells. (1) Upon binding to cell surface receptors, toxins are internalized via multiple pathways. (2) A large portion of toxins is either exocytosed or degraded in the lysosome. (3) A small portion of toxins go through cell and reach ER. (4) Toxin A chain is repelled from ER into cytosol.

1.6 Catalytic mechanism of RTA

After the enzymatically active A chain gets into the cytosol, ricin and shiga toxin start inactivating ribosomes in an extremely efficient fashion. A single ricin A-chain is able to inactivate hundreds of ribosomes per minute (Sandvig and van Deurs 1996). Ricin and shiga toxins are N-glycosidases; they remove a crucial adenine from a conserved rRTA sequence in 28S (a K_m value of about 0.4 μM and a K_{cat} value about 1777 min^{-1}) and the prokaryotic 23S large ribosome and thereby inactivate the ribosome (Endo and Tsurugi 1987).

The specific substrate target of RIPs is termed the sarcin/ricin domain, abbreviated SRD (it is also part of catalytic target of an endoribonuclease, α -Sarcin). This domain is part of the recognition/binding site for translation elongation factor I and II complexes (Moazed, Robertson et al. 1988), and therefore elongation factors cannot interact with the modified subunit, terminating protein synthesis (Liebman, Chernoff et al. 1995). SRD contains the invariant tetra-loop sequence $G_1A_2G_3A_4$, where A_2 is the base to be removed by RIPS; it corresponds to A4324 in rat and A2660 in *E. coli*. The NMR structure of a synthetic oligoribonucleotide (29-mer) that mimics the mammalian SRD showed an irregular helix with a bulged G cross strand A stack and a GAGA tetraloop (Fig. 6a) (Szewczak and Moore 1995). X-ray structures for the mammalian SRD (Correll, Munishkin et al. 1998) and the *E. coli* SRD (Correll, Wool et al. 1999) have confirmed this NMR structure. Orita's high resolution NMR structures (Orita, Nishikawa et al. 1996) for a 12-mer synthetic SRD, together with the other structures reported, allows the absolute requirements for the substrate recognition and catalysis to be depicted. Specifically, in the tetraloop region, G_1 is paired with A_4 while A_2 , G_3 and A_4 form a continuous stack, which was proposed

to help ricin readily access and attack A₂. Although the three-dimensional structure of the GAAA and GCAA substitutions are very similar to the structure of GAGA loop, the former two constructs are not recognized by ricin, suggesting that A₂ and G₃ are required for ricin specific recognition (Orita, Nishikawa et al. 1993; Orita, Nishikawa et al. 1996). Comparison between rat and *E. coli* SRD structures reveal remarkable similarity, however, the “flexible region”, which connects the tetraloop to the A-form helix stem varies. One will see that significant differences in that region shift the conserved sequence domains to the stem. This shift results in 4.9 Å and 7.9 Å deviations for A₂ and G₃ in rat and in *E. coli* respectively, if the two structures are superimposed on their stem parts (Fig. 6b). These structural differences may explain the species specificity for some RIPs. For example, ricin only inactivates eukaryotic ribosomes, whereas shiga toxins can attack both eukaryotic and prokaryotic ribosomes (Suh, Hovde et al. 1998).

Attempts to obtain structural information of RTA complexed with SRD have so far been unsuccessful (Holmberg and Nygard 1996). However, the interactions between the protein and the ligand can be inferred from the X-ray structures of RTA complexed with the substrate analogs, like RTA/FMP (Monzingo, Collins et al. 1993) and RTA/ApG (Monzingo and Robertus 1992). These structures show that, upon binding, Tyr 80 moves away from its native conformation so that the formycin and adenine rings bind in the catalytic site of RTA by aromatic stacking with Tyr 80 and Tyr 123. In addition, a few key active site residues, Val 81, Gly 121, and Arg 180, form specific hydrogen bonds with the polar atoms on the ring of the ligands as indicated in Figure 7. The binding information provided by the structures is also confirmed by site-directed

mutagenesis (Ready, Kim et al. 1991; Kim, Mlsna et al. 1992; Kim and Robertus 1992). Together with kinetic data (Kim, Mlsna et al. 1992; Kim and Robertus 1992), a ricin catalytic mechanism model has been proposed by the Robertus group (Monzingo and Robertus 1992) (Fig. 8). In brief, Arg 180 side chain protonates N3 of the adenine base, which pulls electrons away from the ribose to facilitate cleavage of the N9-ribose C1' bond. This electron movement will leave cation character on the ribose, forming an oxycarbenium transition state, which is stabilized by negatively charged Glu 177. Arg 180 attacks a water molecule in the active site to create the nucleophile, which, in turn, attacks the oxycarbenium C1' on the ribose and complete the glycosidic cleavage between the ribose and the adenine base.

This proposed catalytic mechanism may also suitable for shiga toxin since the active site residues and three dimension structures are identical for both shiga toxin and ricin (Fraser, Chernaia et al. 1994).

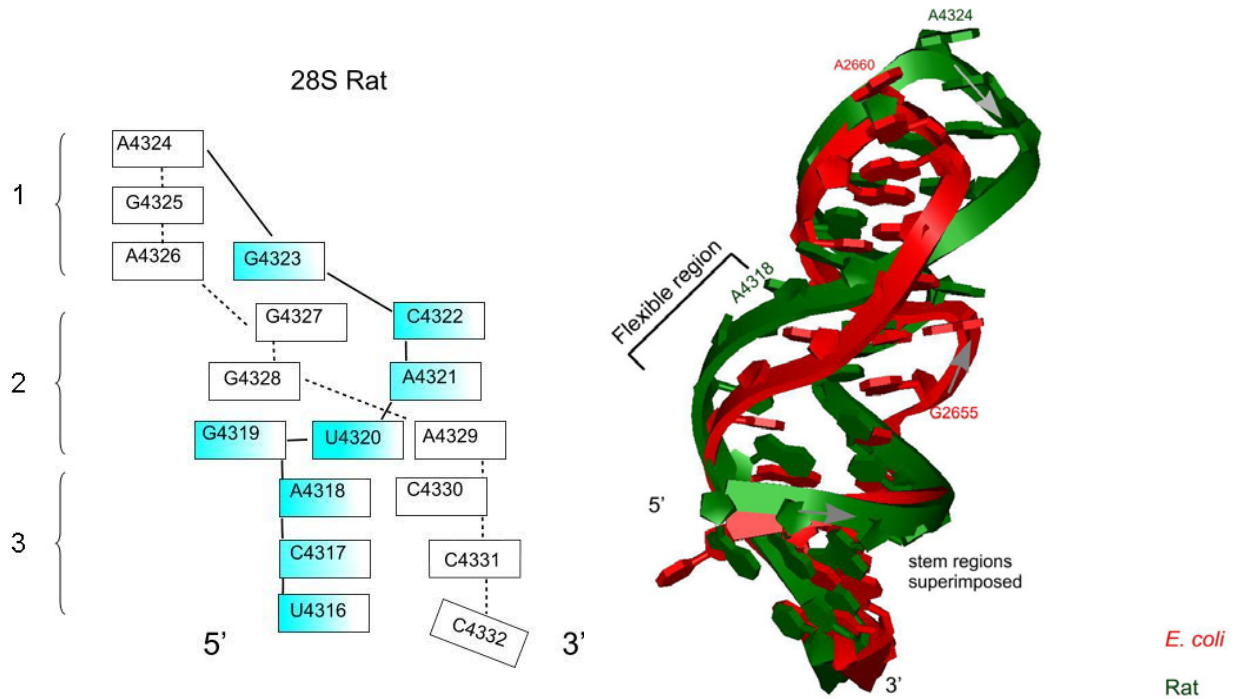


Figure 6 (a). A base-stacking diagram of the SRD (rat). (1). GAGA tetraloop (2).A a Gbulged cross-strand A stack. (3). A flexible region. The residues on 5' side are colored blue. (b). Superimposed stem regions of rat and *E. coli* SRDs. The structure of rat SRD (430D.pdb) is in green, and the structure of *E. coli* SRD (483.pdb) is in red.

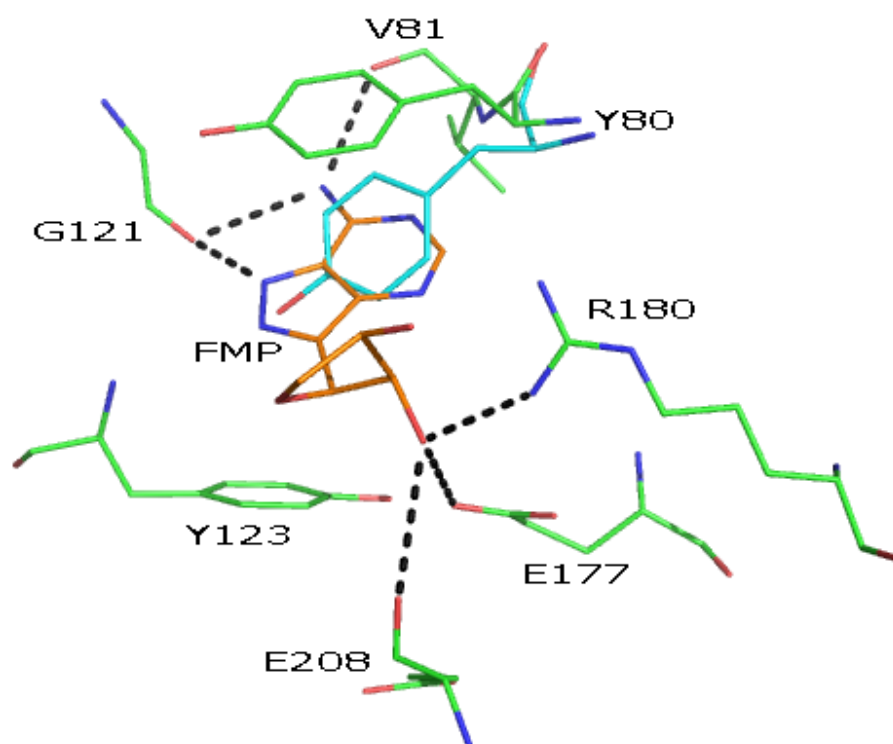


Figure 7. RTA active site upon FMP binding. Y80 is moved away from its native conformation (in cyan) to the “open conformation” (in green). Dotted lines indicate the H-bonds between FMP and RTA active site residues.

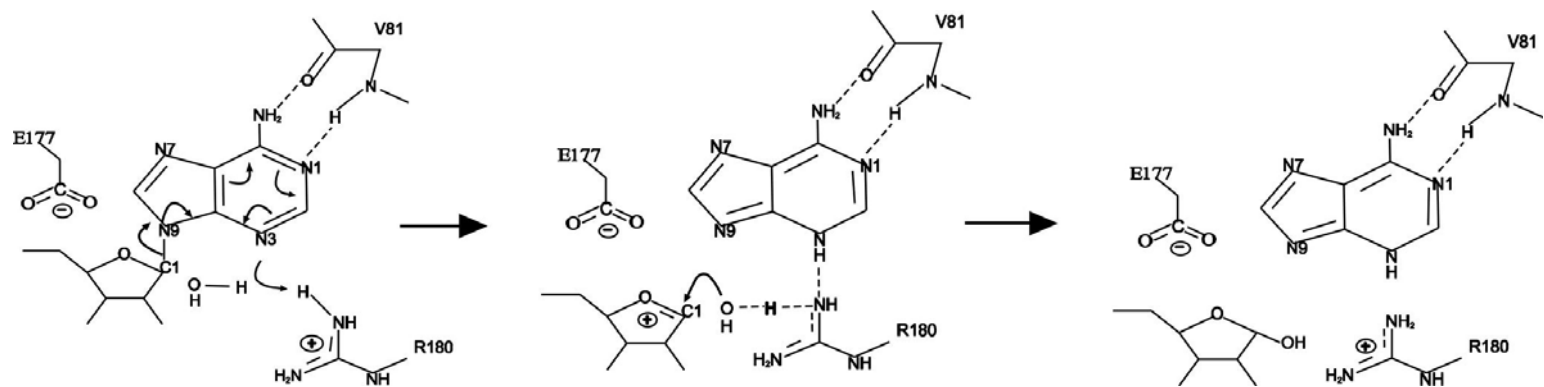


Figure 8. Proposed mechanism for RTA depurination of rRNA

1.7 Ricin inhibitor identification

Ricin inhibitor identification has been focused on the following strategies: (1) eliminate ricin prior to cell binding, (2) block the ricin cell sorting path to prevent it from getting to its site of action, (3) inhibit the enzymatic activity of ricin A chain. Efforts have been made to identify drugs that could interfere with ricin at early intoxication stages (Thompson, Scovill et al. 1995), however no compounds showed a useful effect in protecting animals from exposure to ricin. This may be due to either the compounds cell toxicity or the multiple cellular entry routes available to ricin.

Therefore, most recent attention has been focused on identification of specific compounds that will block the enzymatic activity of ricin A chain. Several classes of ricin A chain inhibitors have been reported so far: (1) small stem-loop oligonucleotides, (2) rationally designed peptides, (3) aptamers selected via screening *in vitro*, and (4) pterin- and purine-like compounds, previously discovered in our laboratory. Schramm's group has reported a number of small stem-loop oligonucleotides which mimic the proposed oxycarbenium ion transition state of the reaction (Chen, Link et al. 1998; Roday, Amukele et al. 2004). They have synthesized strong inhibitors with K_d values in the nanomolar range (Sturm, Roday et al. 2007). However, those inhibitors appear to bind only at low pH, with the maximum inhibition around pH 4. In addition, the stem-loop nucleotides are subject to degradation *in vivo* and may not be suitable for drug development. Small peptides, designed rationally based on the interaction of RTA with its ligands, appear to act as modest inhibitors of ricin (Wang, Feng et al. 2005). *In vitro* aptamer selection generated an RNA ligand for ricin (Hesselberth, Miller et al. 2000). Although the binding for the 31-base aptamer was strong ($K_d \sim 7\text{nM}$), it

was not an effective inhibitor of RTA. This suggests that it might not fully occupy the protein active site.

The Robertus laboratory solved the X-ray structure of ricin (Montfort, Villafranca et al. 1987; Rutenber, Katzin et al. 1991) and defined the mode of substrate binding (Monzingo and Robertus 1992). These structures, together with site directed mutations (Schlossman, Withers et al. 1989; Ready, Kim et al. 1991; Kim, Mlsna et al. 1992), have provided crucial information on the catalytic active site configuration and the mechanism of RTA. This in turn allowed virtual screening and structural-based design to identify small molecule RTA inhibitors. The first inhibitor was pterioic acid, in which the pterin ring bound in the adenine specificity pocket, making a number of key hydrogen bonds (Yan, Hollis et al. 1997). Subsequently they identified adenine and guanine derivatives that bound in a similar fashion (Miller, Schneider et al. 1989); (Yan, Hollis et al. 1997; Miller, Ravikumar et al. 2002) The representative inhibitors for each platform are shown in Figure 9. Most recently, pyrimidine-based RTA inhibitors have been identified and are underway to be further optimized (Bai, Monzingo et al. 2009). This new class of inhibitors works as well as previous purines and pterins, but also has better aqueous solubility than those previous inhibitors; this is important in order to be “drugable”. This discovery and characterization are elaborated in this dissertation.

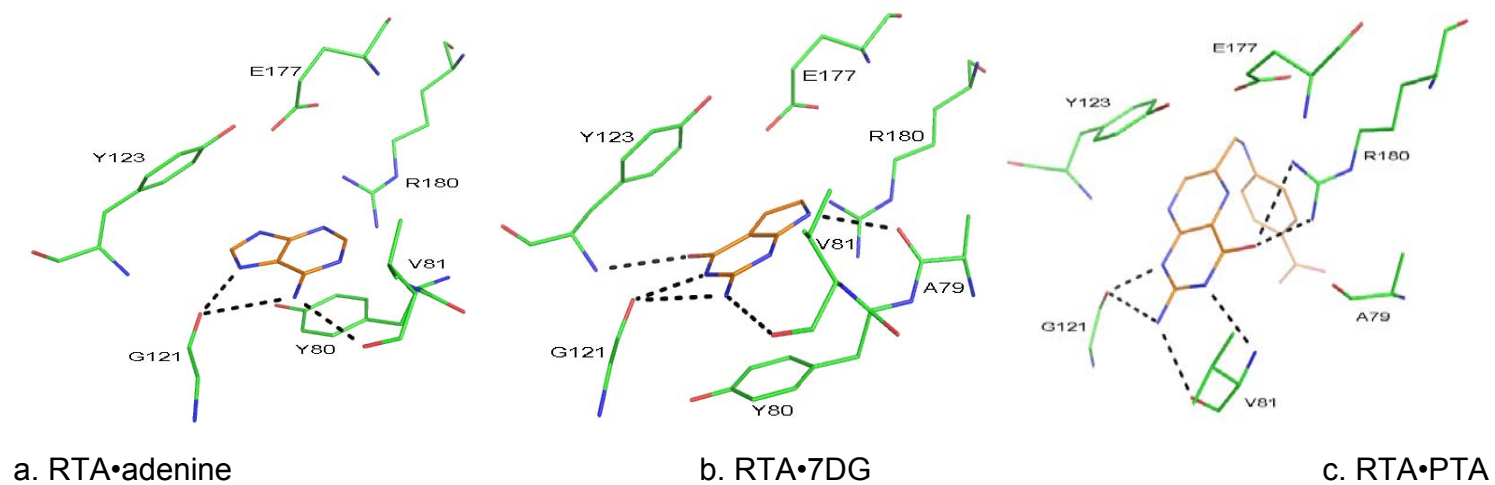


Figure 9. Binding modes for representatives of adenine, guanine, and pterin based RTA inhibitors. Active site residues are in green, ligands are in orange. a. The binding mode of RTA•adenine (1IF3.pdb). b. The binding mode of RTA•7DG (1IL3.pdb) c. The binding mode of RTA•PTA (1BR6.pdb)

1.8 Strategies for RTA inhibitor design and discovery

Methods for discovering new drugs have developed greatly over the past decades, and structure-based drug design (SBDD) has been intensively used (Thomke 1998). This approach requires the three-dimensional structure of a drug target, obtained usually by X-ray crystallography or NMR. The protein targets are structurally displayed by graphics visualization tools. Then scientists can visualize the active site of the target and design specific compounds to fit. In principle, this method facilitated design of drugs in a more rational and efficient way, compared to the older methods in which medicinal chemists systematically modify a chemical precursor and test all synthesized compounds. SBDD has made a number of successful stories in drug discovery, such as drugs for cancer (Appelt, Bacquet et al. 1991), for viral infection (Holmes, Bethell et al. 1993; Lam, Jadhav et al. 1994; Thaisrivongs, Tomich et al. 1994), for heart disease (Patricia C. Weber 1996), for immune disease and cytokines (Glen Spraggon 1996; Michael A. Jarpe 1996), et al.

With the rapid development of computer science, large storage capacity, and fast computation speed Virtual Drug Screening (VDS) has joined SBDD as a viable tool for research. By this method, one can screen many existing compounds *in silico* against the protein target of interest. This process may reduce the range of chemicals that need to be subjected to kinetic tests, and thereby save time and money in drug development. Recently, the increased robustness of computational algorithms and scoring functions, the availability of strong computational power, and timely structural determination of protein targets and their complexes with ligands have made VDS more practical. In the last decade, few drugs have been introduced solely and directly from High throughput

screening (HTS), which is a very expensive process. Many pharmaceutical companies have moved away from large combinational chemistry libraries to more information-rich methods, SBDD (Juan Alvarez 2005).

The ricin structure provides a solid base to use SBDD for ricin inhibitor discovery and development. Indeed, SBDD has been reasonably successful on RTA inhibition identification. The first RTA inhibitor identified by VDS was pteric acid, PTA, with an IC_{50} of 0.6mM. A 2.3 Å crystal structure of the RTA/PTA complex showed that PTA binds the enzyme in a manner consistent with the natural substrate adenine in the active site (Yan, Hollis et al. 1997) (Figure 9). This established the basis for further inhibitor design.

Subsequently, over a dozen purine/pterin-based RTA inhibitors were identified and a few of them were crystallized with RTA (Miller, Ravikumar et al. 2002). Although these make some substrate-like interactions with RTA, the limited solubility and difficulties of derivative synthesis have been obstacles to progress. An alternative RTA inhibitor platform has been identified through VDS recently (Bai, Monzingo et al. 2009). The crystal structure of RTA complexed with a pyrimidine derivative, PBA, showed that the pyrimidine ring has all functional atoms equivalent to those of PTA, and the activity assay also showed that PBA has the same effectiveness as PTA. In addition, PBA is much more soluble than previous RTA inhibitors, which is a benefit for further inhibitor development. The detailed identification and characterization of pyrimidine-based RTA inhibitors are discussed in this dissertation.

Besides, PBA, there are additional potential leads identified through VDS. These include chemicals from the Chembridge Diversity library, which possess totally different structural properties from other known inhibitors. Although the

inhibition mechanisms for these hits have not been characterized yet, the inhibition effects of some hits have been confirmed in cell-based assay. Currently, these hits are underway for crystallization trials.

Using computer simulations, we have modified PBA by elongating its benzoic acid tail in order to occupy a second enzyme pocket (Fig. 10). The second pocket is lined with polar side chains and more deeply set hydrophobic residues, and it is much larger than the specificity pocket. Fragment-based drug discovery could be a feasible tool to search for active binding moieties for this site. Fragment screening is the process of identifying relatively simple and soluble, though often less potent, bioactive molecules. Such fragment hits are usually easy to optimize into clinical candidates with good drug-like properties. Another advantage of fragment screening is that structural information can often be obtained by X-ray crystallography due to their high solubility. Therefore, the process of lead optimization can greatly accelerate based on timely structure information.

A high throughput assay is usually used to test the hits from VDS, since the number of compounds is fairly large. With the recent advances in screening methodologies, liquid-handling technologies, and robotic instrumentation, HTS assay has been made practical for the early stages of drug discovery. However, HTS needs to be designed with careful considerations, such as simplicity, robustness, reproducibility, and cost. In addition, once hits are found and kinetically verified, there should be other assays to confirm whether they are real or not (Coan and Shoichet 2008). Unfortunately, there is still no good HTS assay available for ricin inhibitors. In this dissertation, I report considerable efforts to find a good assay. We have tried LC-HPLC, Circular Dichroism, Fluorescence

Polarization and Isothermal Titration Calorimetry, but found that none of them is perfect in a HTS format. It does appear that ITC would be a very useful assay to confirm, and quantitatively analyze, inhibitor candidates. We also collaborate with Dr. Nicholas Mantis at the Wadsworth Center in NY to screen huge libraries using a cell-based HTS assay. We used an in vitro system to confirm that RTA, and not other aspects of intoxication, was the target of the inhibitor candidates. We have found, so far, 8 hits against RTA and 6 against SltA. The details are discussed in this dissertation

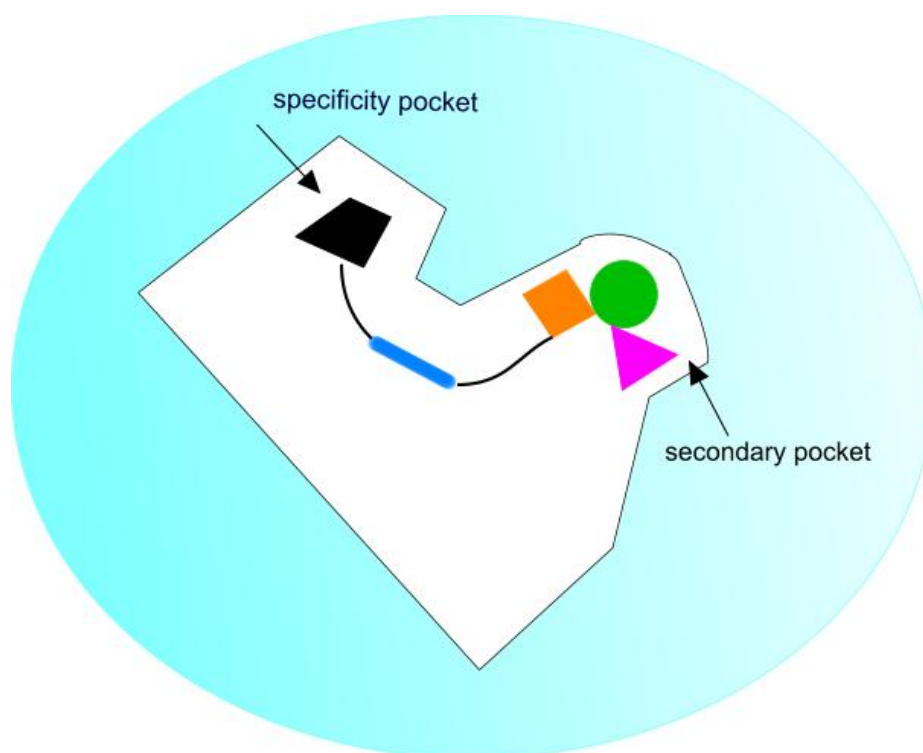


Figure 10. Illustration of the design of potential strong inhibitors. The binding scaffold for the specificity pocket is in black, and the fragments that would bind in the second pocket are in orange, green and pink. The blue rod represents the linker between the two major binding portions.

CHAPTER 2: CONTINUING STRUCTURAL STUDIES ON PURINE AND PTERIN BASED RTA INHIBITORS

2.1 INTRODUCTION

Previous studies on RTA inhibitors have identified more than a dozen purine and pterin based inhibitors (Miller, 2002; Yan, 1997; Yan, 1998). Table 1 lists all RTA inhibitors identified prior to this work. However, all these inhibitors are weak binders with IC_{50} s around 1mM. Therefore we continued work in this direction by collaborating with Dr. Sean Kerwin in the Pharmacy School and Dr. Eric Anslyn in the Department of Chemistry and Biochemistry. Our goal was to find RTA inhibitors that are more potent and more favorable for further drug development. In this work, seven more RTA inhibitors derived from purine and pterin platforms are reported, and three of them are crystallized with RTA as complexes.

2.2 MATERIALS AND METHODS

2.2.1 RTA expression and purification

The expression vector for RTA was constructed from the pUC18 vector and recombinant RTA was overexpressed in *Escherichia coli* JM101 as described previously (Ready, 1991). We also found that expression in the JM105 bacterial strain performed as well as JM101, and the strains were used interchangeably.

Transformed cells were plated on LB/AMP plates and incubated overnight at 37 °C. A single colony was picked for inoculation of 50 ml cultures in 2YT media. Cells were grown overnight in a shaker at 37 °C, after addition of 50 µl of

100mg/ml ampicillin (amp) stock solution. An aliquot of the overnight culture, usually 10 ml, was transferred into 2L flasks (with aerating edges) containing 500 ml of 2YT media with 100 μ g/ml amp to make the starting OD about 0.1. Cells were grown at 37°C for about 2 hours while shaking at 275 rpm until the OD reached 0.5 - 1. Then protein expression was induced with 100 μ M IPTG. Protein expression was allowed to proceed for about 4 hours, and cells were harvested by centrifugation at 5000g in a GSA Sorvall rotor for 10 min. Cells were resuspended in 30 ml to 40 ml phosphate buffer (5 mM NaPO₄, pH 6.5). The cells were lysed in a French pressure cell 2 to 3 times at 1,000 psi and the lysate centrifuged at 35,000 rpm for 1 hour in a 42.1 Ti rotor (Beckman). The supernatant was filtered in a 0.2 μ m syringe filter, loaded on a pre-equilibrated CM sepharose column (Sigma), and washed with 4 L of phosphate buffer (5 mM NaPO₄, pH 6.5) overnight; 2 L wash buffer was sufficient for this step. The column was washed with 0.1 M NaCl in the buffer above for at least 200 ml before gradient elution. The column head was lowered down to about 1 cm above the column bed, and then the elution step started. The elution buffer used was phosphate (5 mM NaPO₄, pH 6.5) with a 0.1 M-0.3 M NaCl gradient (about 500 ml). Protein purity was estimated by SDS-PAGE and pure fractions were combined. The fresh protein was assayed for activity, and a portion of RTA protein was either snap-frozen in 50 μ l aliquots for later assay use, or dialyzed and concentrated to 2 ~ 3 mg/ml for crystallization.

2.2.2 Determination of RTA and RTA inhibitor activity

The activity of various RTA inhibitors was tested by an *in vitro* translation assay using rabbit reticulocyte lysate systems (Ambion/Promega). We used two

systems which differ in how the protein synthesis reporter is incorporated, and how the protein synthesis level is measured.

One system used ^{35}S labeled methionine to label the protein translated from the endogenous hemoglobin mRNA contained in untreated rabbit reticulocyte lysate (Ambion). A concentrated "Super mix" was prepared first : 4 μl of translation mix (38.6 mg/ml creatine phosphate in Tris•HCl pH 7.6, 10 mg/ml creatine kinase Tris•HCl pH 7.6, 1mM spermidine, 10mM ATP, 2mM GTP, 1.25M potassium acetate, 625 mM sodium chloride) (Sigma-Aldrich), 2 μl of ^{35}S -methionine (10 $\mu\text{Ci}/\mu\text{l}$ stock) (Amersham), 1.5 μl of magnesium acetate (Sigma-Aldrich) (25 mM stock) and 1 μl of amino acid mix (Ambion). The enzyme reaction was then set up in the following order: 1) 10 μl of pre-mixed RTA/compounds (5 min pre-incubation) was added in the 20 μl of lysate, followed by 2) a 6 min incubation at room temperature, then, 3) 8.5 μl of super mix was added to initiate the translation reaction, 4) protein synthesis was allowed to proceed for 60 min in a 30 °C water bath, 5) 20 μl of reaction mix was removed for quantitative analysis. The reaction mixture was decolorized by addition of 0.5 ml NaOH (1N), and then precipitated by addition of 1 ml of cold 25% TCA (w/v). After 5 min incubation on ice, the precipitants were collected by vacuum filtration (Whatman GF/A glass fiber filter). Filters were rinsed with 2 ml of 10 % TCA, and 10 ml of 1% TCA, and dried by vacuum for 1 min. Filters were oven-dried for 10 min at 60 °C, and then were placed into scintillation vials with 5 ml Econo-Safe (biodegradable) counting cocktail, and counted in Beckman liquid scintillation counter to measure the incorporation of isotope, which was proportional to the protein synthesis level.

Another assay system used Rabbit Reticulocyte Lysate (Promega) that was treated with micrococcal nuclease to destroy endogenous mRNA. This reduced background translation to a minimum. This system allowed us to introduce mRNA for a signal protein which can be detected after translation (Pelham, 1976). In this case, luciferase mRNA was used to synthesize luciferase protein, which can generate light in the presence of its substrate and ATP. The *in vitro* translation assay was scaled down from the original protocol: 7 μ l rabbit reticulocyte lysate, 0.2 μ l amino acid complete, 0.2 μ l RNasin ribonuclease inhibitor, 0.4 μ l luciferase control RNA (20 ng/ μ l) for each reaction sample. After adding the RTA and various concentrations of compounds, translation was initiated by adding translation mix (amino acids, RNasin ribonuclease inhibitor, luciferase control RNA, nuclease-free water), followed by a 90 min incubation at 30°C. The reaction was stopped by simply freezing the reaction tube at -20 °C. Finally, 2.5 μ l of reaction mix was mixed with 50 μ l of luciferase substrate reagent (Promega), and the luminescence was immediately measured on a Perkin Elmer Envision luminometer (Waltham MA). Figure 11 illustrates how the luciferase assay works.

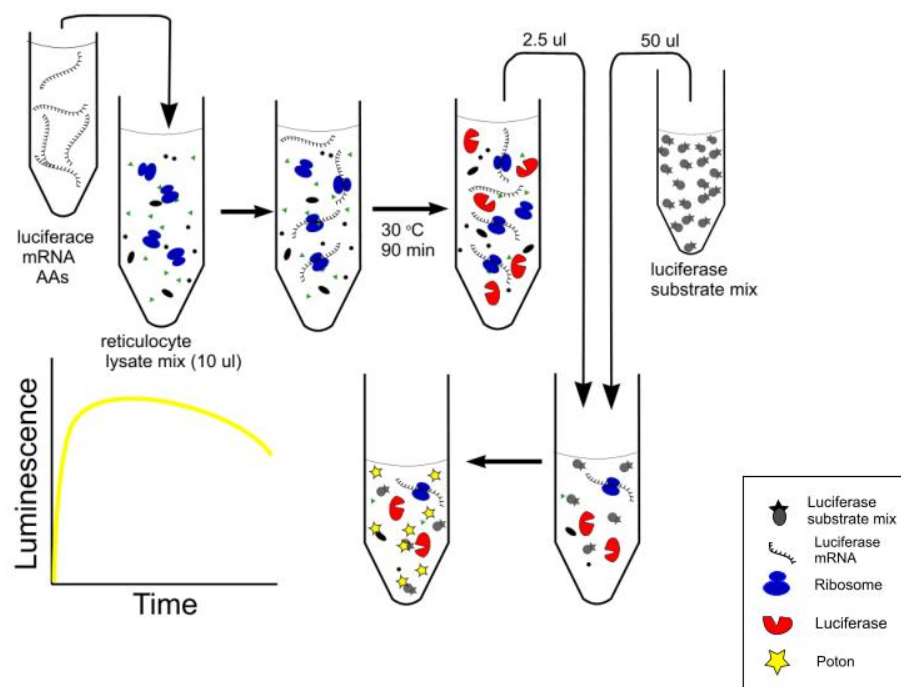


Figure 11. The cartoon of the luciferase assay.

For both assay systems RTA dose response were carried out to determine the RTA concentration that was going to be used in the inhibition assay. The concentration of RTA causing 95 % inhibition of protein synthesis was used in the subsequent RTA compound inhibition assays; this was usually about 0.2 pM. All the compounds tested were dissolved in 0.05 N KOH, prior to addition to the RTA-containing solution.

2.2.3 Crystallization of RTA

RTA was dialyzed into a 75 mM Tris pH 8.9, 1mM EDTA solution using dialysis tubing with a 6-8,000 MW cutoff. The protein was concentrated to 2.0~3.0 mg/ml in an Amicon concentrator using Millipore YM10 filter discs. In order to help preserve RTA, 10 mM BME could be added into concentrated RTA solution. Monoclinic RTA crystal were obtained by seeding at the condition described previously (Robertus, 1987; Mlsna, 1993). Briefly, 20 μ l of protein at 2-3 mg/ml in mother liquor (75 mM Tris-HCl, pH 8.9, 10 mM BME, 1 mM EDTA, 4.1% PEG MW 8000) was suspended over 200 μ l of mother liquor reservoir at 4°C.

To form complexes, the Ligand was soaked into RTA crystals, with an estimated final ligand concentration about 1 to 5 mM. After a short soaking time period, typically overnight, the crystal was mounted for X-ray data collection. The crystal was transferred into a drop of artificial mother liquor containing 10% glycerol for 2 minutes, and then transferred to paratone oil. After removing surface-bound water the crystal was mounted in a cryoloop (Hampton Research) and flash-frozen in liquid nitrogen prior to mounting in a nitrogen cold stream.

2.2.4 Inhibitor soaking and co-crystallization

The compounds 7-methyl-9-deazaguanine (7M9DG) was synthesized in Dr. Kerwin's laboratory and 7-carboxy pterin (7CP) was synthesized in Dr. Anslyn's laboratory. Each was dissolved in 0.05 N NaOH; the stock concentrations for these ligands were 50 mM and 60 mM respectively. 1 μ l of ligand solution was added into a 20 μ l RTA crystal drop for soaking. The final concentration of ligands was estimated at 2.5 mM, and 3mM respectively.

The compound 7-proporgyl-9-deazaguanine (7P9DG) was synthesized in the Kerwin laboratory, It was co-crystallized with 2.3 mg/ml RTA in 75 mM Tris-HCl pH 8.25, 10 mM BME, 1 mM EDTA, 4.1% PEG MW 8000 with 3% ethanol (v/v).

2.2.5 X-ray data collection and structure determination

Diffraction data for RTA•7M9DG and RTA•7P9DG were collected on a Rigaku Raxis IV area detector. X-rays were generated by a Rigaku RU200 rotating anode generator operated at 50 mV, 100 mA. The data for RTA•7CP were collected at 100 K on a MAR345 image plate detector (Marresearch) with X-rays generated by a Rigaku MicroMax007 rotating anode generator operated at 40 mV, 30 mA. Diffraction images were processed, data reduced, merged, and scaled, using the program HKL2000 (Otwinowski, 1997).

The first model, in each case, was generated by molecular replacement, using the MOLREP routine in CCP4 (CCP4, 1994). The starting model was the monoclinic form of recombinant RTA (PDB code 1RTC) (Mlsna, 1993). The coordinates of 1RTC were modified by mutating Tyr80 to Ala in order to avoid bias in electron density map toward the closed form seen in the apo protein. Further refinements and electron density map generation were carried out using the CCP4 routine REFMAC (Murshudov, 1997), and the CNS suite of programs

(Brunger, 1998). The program Coot (Emsley, 2004) was used for manual model building, water molecule addition, and visualization.

The procedure of ligand incorporation into crystallographic refinement is as follows. First, one can get a ligand coordinate file, topology file, and parameter file from the Dundee ProdrG2 Server (Schüttelkopf, 2004). As the names imply, these files define the atom names, bond lengths and angles, and torsional freedom for the ligand molecules. These files are required for restrained refinement in CNS. When REFMAC, in CCP4, was used, "ligand.cif" file was required for modeling and refinement. These ligand files complement the internal parameter files each program maintains for protein structure, and together they allow the atoms of the complex to be moved with respect to one another, while preserving the bonding patterns expected for the protein and the small molecule.

For visualization, and manual rebuilding, the ligand.pdb, and ligand.cif were read into COOT. After activating the "Calculate/Rotate/Translate Zone" tool in COOT, the ligand can be moved into its electron density. Once the ligand was well-placed in the density, it was merged to the protein coordinate file and saved. Then, this saved coordinate file was read into REFMAC for a new run of refinement.

2.3 Results and discussion

Over 50 compounds derived from the purine and pterin platforms were tested for RTA inhibition in the activity assay. These were acquired either from the Dr. Kerwin or Dr. Anslyn's laboratory, or were purchased from commercial sources. Seven compounds were confirmed by the activity assay as RTA inhibitors (to be described later) and were soaked into, or co-crystallized with,

RTA (Table 1). Of these, three formed stable complexes. The IC_{50} values of these inhibitors were between 0.3 mM and 4.7 mM. They are clearly very weak and do not show much improvement in binding affinity compared with previous RTA inhibitors derived from these platforms. However, the X-ray structures do provide some interesting structure information, and they will be discussed in turn. The inhibitors 7M9DG and 7P9DG are derivatives of 9DG, for which an X-ray complex was already known (Miller, 2002). These derivatives added a methyl and propargyl group respectively to N7. The IC_{50} s for 9DG, 7M9DG, and 7P9DG are 1.8, 1.8, and 2.1 mM respectively. That is, the minor modifications had little effect on the apparent binding affinity. An electron density map for 7M9DG is shown in Figure 13 and that for 7P9DG in Figure 12b.

. Figure 14 shows the superimposed X-ray structures of RTA•7M9DG, RTA•7HP9DG, RTA•9DG, and RTA•7DG a compound analyzed previously (Miller et al, 2002). The binding poses for RTA•7M9DG and RTA•7HP9DG are essentially the same, and they strongly resemble the pose for 7DG. In each case, the exocyclic amine donates hydrogen-bonds to the carbonyl oxygen of V81 and G121. N3 is protonated and donates a hydrogen-bond to carbonyl oxygen of G121, and the carbonyl oxygen on the C₆ position receives a hydrogen-bond from the R180 side chain. Interestingly, the binding mode of RTA•9DG differs. The guanine ring of 9DG is flipped when compared with the corresponding groups of 7M9DG, 7P9DG, and 7DG. Specifically, the carbonyl oxygen on the C₆ position of 9DG forms hydrogen-bond with the amido group of Y123. This observation has been commented on previously based on a comparison of 7DG and 9DG (Miller et al, 2002). The structures suggest that (1) the platform ring of these ligands differ very slightly in interaction energy and

could be rotated to accommodate different branching groups, (2) a single pyrimidine ring, with appropriate functional groups, should function the same as guanine in terms of binding in the specific pocket.

Another interesting observation was that the ligand 7-propargyl-9-deazaguanine (7P9DG) was altered during crystallization; it appears that water added across the triple bond to form 7-(2-hydroxypropene)-9-deazaguanine (7HP9DG). The explanation might be that the ethyl group is hydrolyzed (Markownikoff, 1870) to the enol group at crystallization condition, which can also be tautomerized to the keto group. In aqueous solution, the keto form predominates over the enol form at equilibrium. Nonetheless, the enol form is very important for some reactions. The more stable keto form was built into the difference electron density map in the beginning of refinements. However, the map indicated that the carbonyl oxygen should be put in a position where it would be chemically repelled by neighboring protein oxygen atoms on the side chain of E177 and the backbone oxygen of E208. If the keto group was replaced by the enol, the R_{free} and R_{work} for the refinement dropped by about 2% after several run of refinements. More importantly, in this tautomer, the enol group would be hydrogen-bonded with the E208 backbone oxygen. This favorable interaction may stabilize the enol form. Figure 12a shows the conversion of 7P9DG, and 12b illustrates electron density map for RTA/7HP9DG complex.

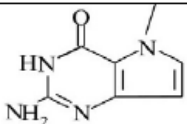
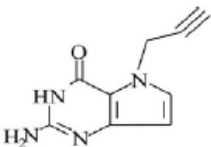
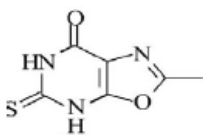
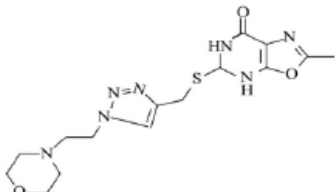
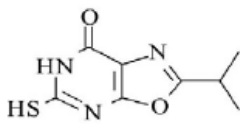
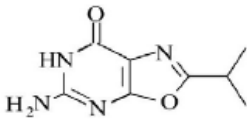
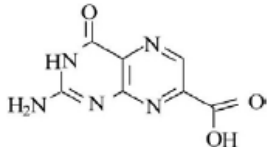
	2D Structure	Name	IC ₅₀	Resolution (Å)
1		KeRCN 088 (7M9DG)	2.2mM	1.8
2		SM91 (7P9DG)	1.6mM	2.1
3		WT063	4.7mM	N/A
4		KeRCN120	1.6mM	N/A
5		KeRCN127	0.27mM(IC ₃₀) Inhibit the protein synthesis above 2.5mM	N/A
6		KeRCN128	3.86mM(IC ₂₃) Inhibit the protein synthesis above 4mM	N/A
7		7CP	0.3mM	2.0

Table 1. Seven RTA inhibitors derived from purine and pterin families. These inhibitors followed the initial success with pteric acid.

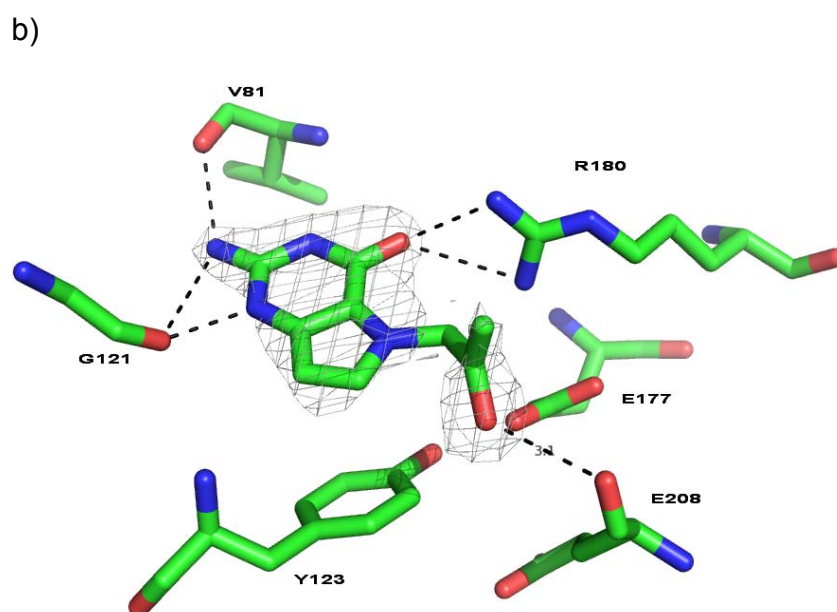
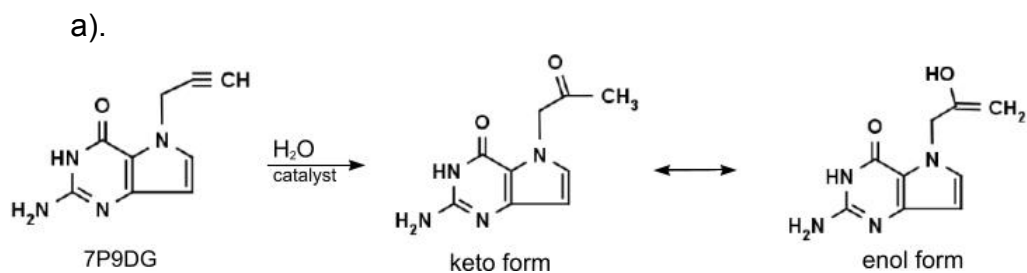


Figure 12 (a) Hydration and conversion of 7P9DG. (b) The active site of RTA•7H*9DG crystal complex. The electron density map for the ligand is in mesh; this is the omit map of the form Fo-Fc, where Fc reflects the calculated Fs in the absence of ligand..

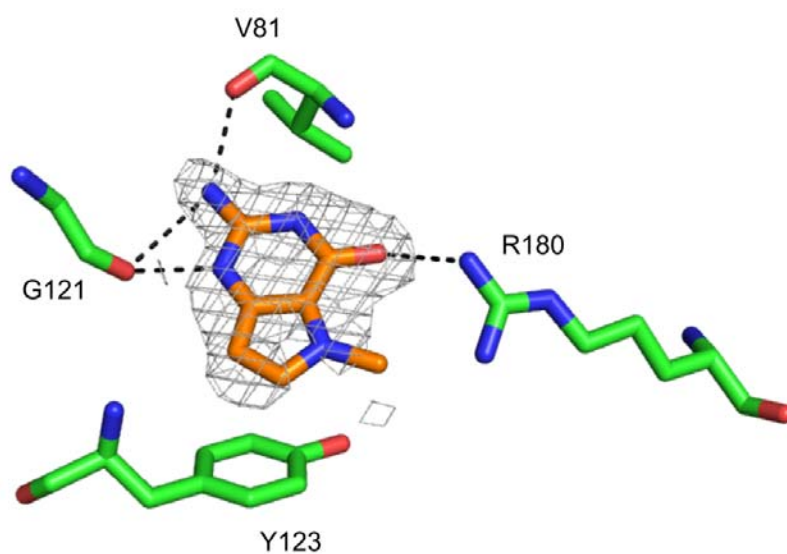


Figure 13. The active site of RTA•7M9DG crystal complex. The electron density map for the ligand is in mesh; this is the omit map of the form $F_o - F_c$, where F_c reflects the calculated F_s in the absence of ligand.

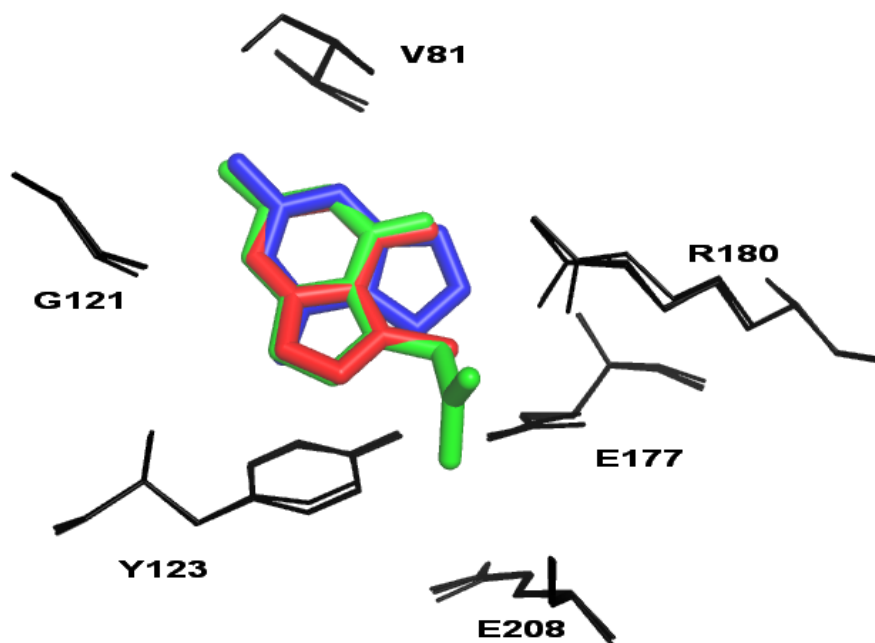


Figure 14. Superposition of RTA•9DG, RTA•7M9DG/ and RTA•7HP9DG. The protein said chains are in black lines, ligands in stickers with difference color. 7DG is in blue, 7M9DG in red, and 7HP9DG in green.

In addition to the guanine based inhibitors, several pterin based RTA inhibitors have been previously characterized, including pteric acid (PTA) and neopterin (Yan et al, 1997). For this project I also solved the structure of a novel pterin based inhibitor, 7-carboxy pterin (7CP). This compound was synthesized in the Anslyn laboratory as part of an ongoing project to add large pendants to the established pterin platform. An omit electron density map, contoured around the inhibitor, is shown in Figure 15.

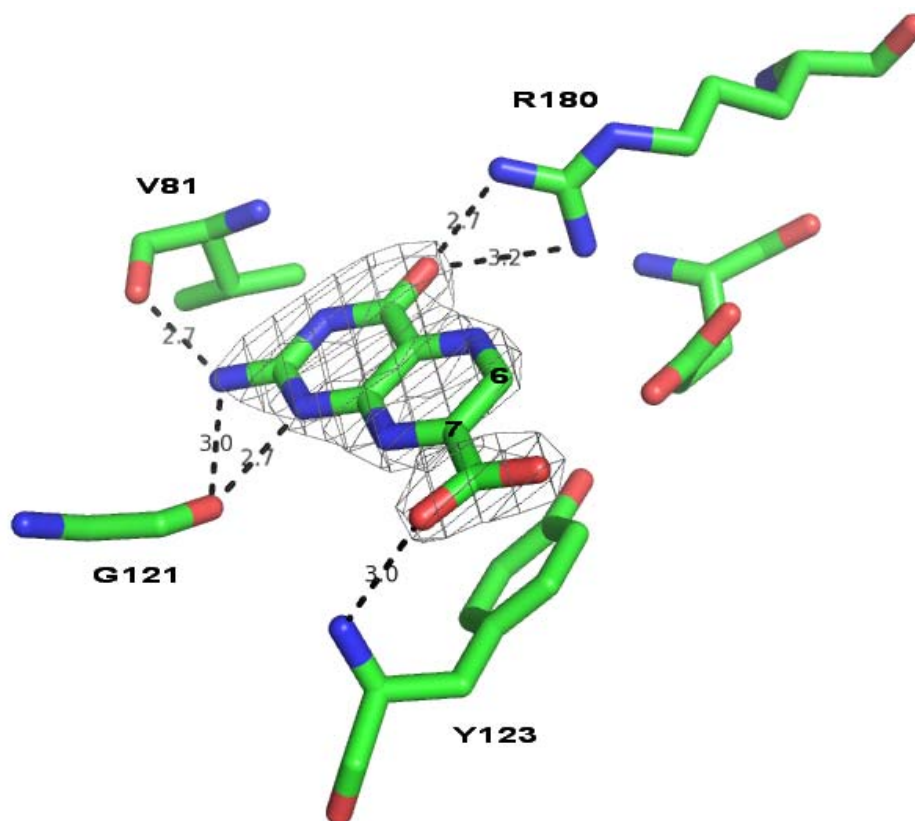


Figure 15. X-ray structure of RTA•7CP. The electron density is an omit difference map. The hydrogen bonds are indicated with dotted lines and distances between the hydrogen-bonding atoms are label by numbers with unit Å.

Like PTA and neopterin, the exocyclic amine of 7CP hydrogen-bonds with carbonyl oxygens of V81 and G121, N3 was protonated and donates a hydrogen-bond to carbonyl oxygen of G121, and the carbonyl oxygen on the C₆ position hydrogen-bonds with the R180 side chain. The carboxylate pendant at position 7 is much shorter than those for PTA or neopterin and the oxygen on the carboxylate tail is hydrogen-bonding to the amido group of Y123. We also tested 6 carboxy-pterin in the activity assay, however it did not show any inhibition. Figure 14 indicated that, given a similar orientation of the pterin, there would be repulsion between the ligand and RTA if the carboxylate group was linked with C₆; the negatively charged E177 side chain would be too close in distance to the carboxylate group on C₆ position.

Compound 2-methyl-5-[1-(2-morpholin-4-yl-ethyl)-1H-[1,2,3]triazol-4-ylmethylsufanyl-oxazolo[5,4-d]pyrimidin-7(6H)-one (KeCRN120) was the only working compounds that was a derivate from guanine on the C₂ position (Compound 4 in Table1). The IC₅₀ of this compound was 1.6mM. The compound substituted the critical amine group on C₂ position, which was proposed to be required for specific pocket binding. It is hard to explain how such a replacement would not jeopardize the inhibitory effect of the compound on RTA. We tried to crystallize the compound with RTA in a complexed form, which could reveal the inhibition mechanism, however, these efforts failed.

Based on these experiments and results, we provided the structure and activity relationship analysis (SAR), shown in Table 2. For guanine derivatives, most modifications were made on guanine C₇ and C₈ positions; modifications on C₉ seemed more favorable, perhaps due to steric effects. In addition, the branching groups at this position are more diverse. For example, the length of the

pendants could be 1~8 atoms long, and could be either linear or branched, allowing a wider range of interactions. However, our experiments showed that an alkane derivative ending with an amine group should be avoided; these modifications rendered the ligand toxic to protein synthesis. Not many modifications have been made for the adenine platform since the adenine is relatively weaker binder compared with the other two platforms. For pterins, modifications were made on both C₆ and C₇ positions. As discussed above, C₆ is very close to E177, so no short acids moieties could be build on this position. However, this problem would be overcome if a flexible nonacidic linker was used to link other functional groups at this position, like the known inhibitor PTA. However, RTA inhibitors derived to date from these platforms were all weak, and solubility was poor. As a consequence, we went to different strategies for our RTA inhibitor search.

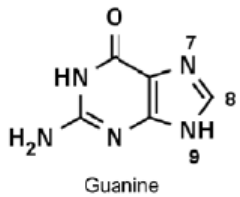
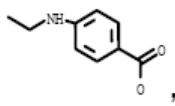
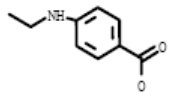
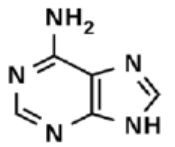
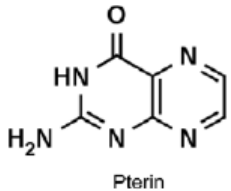
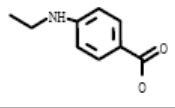
Platform		Pendants	Effect
 Guanine	7	 -CH ₃ ,	+
		-COOH	-
	8	 -CH ₃ , -CH(CH ₃) ₂ , -SH,	+
		-(CH ₂) _n NH ₂ (n=2,3,4) (inhibit protein synthesis)	-
 Adenine	8	-NH(CH ₂) _n NH ₂ , (inhibit protein synthesis)	-
 Pterin	6	 -COOH, -CH ₃	+
		-COOH, -CH ₃	-
	7	-COOH,	+
		-C(CH ₃) ₃ ,	-

Table 2. SAR analysis for purine/pterin-based RTA inhibitors.

Chapter 3: RTA INHIBITORS IDENTIFIED BY VIRTUAL DRUG SCREENING

3.1 INTRODUCTION

Virtual drug screening, VDS, has been recognized as a new tool for drug discovery since a number of successful stories in this field were reported. (Shoichet, McGovern et al. 2002). Today it is common to incorporate VDS into initial drug lead searches, provided the information required for VDS is available. In particular, reliable structural information for the drug target is required for VDS. Given a three-dimensional protein model, computer programs can predict a ligand binding site in that target, and potential ligands from a list of small molecule structures can be fit into the pocket by the process called “docking”. This process searches over many possible interactions between a potential ligand and the receptor, and then ranks the binding “poses” estimated binding energy using a scoring function algorithm. For the ricin project, the structures of both the apo protein and complexes of the protein with known inhibitors were solved in our laboratory, and therefore VDS would be an appropriate method for novel inhibitor discovery and development.

It has been decades since scientists started developing docking programs by employing computation technology. Originally, docking programs were rigid, that is both the target and potential ligand were treated as rigid bodies that can not change their internal spatial arrangement during the docking process. Clearly, the search space was very limited, and given that a ligand generally has many configurations of similar internal energy, and that the target protein may be flexible, such dockings were not terribly useful. Today, we can benefit from the

development of more powerful algorithms in computation biology, and all modern docking programs allow bond rotation flexibility of ligands during the docking process. Furthermore, in recent years, increasing efforts have been made to incorporate protein flexibility into these docking programs, at least for the side chains of active site residues or loops in the proximity of active site of protein targets (McMartin and Bohacek 1997; Schneider and Bohm 2002; Schulz-Gasch and Stahl 2003; Warren, Andrews et al. 2006).

The most commonly used docking programs today are probably FlexX (Rarey, Kramer et al. 1996), DOCK (Ewing, Makino et al. 2001), GOLD (Jones, Willett et al. 1997), ICM (Ruben Abagyan 1994), and Glide (Richard A. Friesner 2004) among over a dozen docking programs. These programs use different docking strategies and different scoring algorithms (Schneider and Bohm 2002). According to recent evaluations by pharmaceutical laboratories on these docking programs (Perola, Walters et al. 2004; Warren, Andrews et al. 2006), all of the docking programs were able to predict ligand conformations similar to what had been observed crystallographically for many protein targets. However, the scoring functions were less successful at ranking the results – which pose is “better” than the others and how to correlate the binding pose and actual binding affinity. Another observation was that no single program performed well for all of the targets. These problems have attracted much attention and effort recently, and some progress has been made by using different strategies, such as ‘consensus scoring’ (Charifson, Corkery et al. 1999), training scoring functions for specific protein targets (Koehler and Villar 2000), and combining multi-point pharmacophore searching with docking scores reflecting the constraints of the receptor binding site (Mason and Cheney 2000). Although VDS has not been

ideal, it has generated a number of success stories in drug discovery and development; among the most exciting may be the design of the neuraminidase inhibitor zanamivir. Under the guidance of computer simulations, the simple replacement of a hydroxyl group with a positively charged guanidinium group, increased the activity of the ligand by four orders of magnitude (von Itzstein, Wu et al. 1993).

For our search program, we employed VDS in order to find novel RTA inhibitors. The docking programs we used are GOLD (Jones, Willett et al. 1997), ICM (Ruben Abagyan 1994), Surflex (Jain 2003), and eHiTs (Zsoldos, Reid et al. 2006; Zsoldos, Reid et al. 2007). As previously discussed, all these programs can make good predictions of the correct binding poses, but their scoring functions may not be predictive of true binding strength. However, we have observed a promising phenomenon - that running multiple programs for one docking task may increase the hits rate. This will be discussed in this chapter in detail. We screened several compound libraries and will report the results primarily from two large compound libraries, the Sigma Aldrich Diversity Library and the Chembridge Library, which includes the Chembridge Diversity Library, Chembridge Kinase Library, and Chembridge Fragment Library. We will report that the most useful programs for these screens appears to be ICM (Ruben Abagyan 1994) and GOLD. We have identified at least three new RTA inhibitor platforms, and obtained X-ray structures of their complexes with RTA.

3.2 MATERIALS AND METHODS

3.2.1 Compound database

The Sigma Aldrich Diversity Library contains 48,500 compounds, listed as a concatenated sdf file (that is containing three dimensional coordinates and other chemical properties of the compound). The Chembridge Diversity Library contains 49797 compounds, with 4541 in Chembridge Fragment Library, and 12689 in Chembridge Kinase Library. The libraries were generally downloaded in a two dimensional form from corresponding vendor web sites and converted to a 3D SDF format by program the ICM (Internal Coordinate Mechanics). If necessarily, other ligand formats were converted to SDF using Open Babel. The molecular visualization system, Pymol (Delano Scientific LLC), was used to display the docking input and output files for presentations.

3.2.2 Receptor file preparation

The X-ray structure of ricin A chain, RTA, shows the adenine substrate binding pocket is blocked, or closed, by the side chain of Tyr 80 (Mlsna et al, 1993). Its side chain is displaced into an "open" form when substrate analogs or inhibitors are bound (Monzingo et al, 1992; Yan et al 1997). Docking requires that RTA be in the open configuration. To do this, we used the protein portion of the RTA-neopterin complex (pdb code 1BR5) (Yan, Hollis et al. 1997), from which the ligand had been removed. For docking in ICM, the binding site was identified by the program itself. For docking in GOLD, the binding site was defined by GOLD using the assigned ligand neopterin.

For RTA docking in the native closed conformation, the PDB file of recombinant RTA, 1RTC.pdb (Mlsna, Monzingo et al. 1993) was used. The binding site was assigned by ICM in ICM docking by cavity search; in GOLD, the

binding site was defined by assigning “active site atoms” with a radius of 10 Å of a user specified point.

3.2.3 Computer workstation

Virtual screening was carried out by parallel processing on the TI3D Drug Discovery Cluster. The cluster is equipped with 16 HP Proliant BL35P blade servers, each of which has two dual core AMD Opteron 2.4GHz processors for a total of 64 processors. Each blade contains 8 GB of memory and a 6 GB ATA hard disk drive. The front-end of the cluster is an HP xw9300 workstation with a dual core AMD Opteron 2.4GHz processor, an NVIDIA Quadro Fx4500 graphic card, 4 GB memory, and two 500 GB SATA hard drives. An HP Proliant DL380 G4 storage server is attached with an Intel Xeon Processor (3.4 GHz), which supports a RAID5 network attached storage system with five 500 GB SATA disc drives.

3.2.4 Docking with ICM

The screening method for ICM has been described previously. (Abagyan and Totrov 2001) Basically, ICM uses a Monte Carlo global energy optimization to dock flexible ligands into a defined active site of the receptor, which was set up as a rigid molecule. Large-scale random movement of the ligand is combined with a gradient local minimization at every step along the whole procedure. Finally, a small set of “binders” are discriminated from non-binders by the scoring function. In particular for the RTA inhibitor search, the coordinates of RTA with all hydrogen atoms added (either open conformation or closed conformation) were read into ICM and converted to an ICM object by the program’s internal energy optimization process. A specific object name and receptor needed to be set for each docking task. During the active site search, which is one of steps in receptor

set-up, two main cavities were found and listed on a table. The first cavity was the actual specificity pocket known from previous work to bind inhibitors, and second cavity was the “secondary pocket” on the opposite face of the Y80 side chain from the specificity pocket. For the RTA open conformation docking, the first cavity was selected, and the box size that defines the mapping area was then further defined by user. The compound database, that is the concatenated SDF library file, was converted into an indexed file for batch docking. Before launching an ICM run, a scoring threshold was set up as +32, which determines how many compounds from the screen will be listed, base on ICM scores. The score is an indicator of the free energy of binding so the more negative scores reflect better predicted binding. The program default threshold was -32, which indicates a strong potential binder. Here, we set the threshold as +32 in order to make sure that the whole library can be docked and scored in the output file.

3.2.5 Docking with GOLD

GOLD (Genetic Optimization for Ligand Docking) is a program used to predict how flexible molecules bind to proteins, using a non-deterministic sampling Genetic Algorithm (GA) (Jones, Willett et al. 1995; Jones, Willett et al. 1997). For docking RTA in the open conformation, the receptor was an RTA/neopterin complex structure (1BR5.pdb), from which the ligand coordinates were removed. All hydrogens were added to the protein before reading into GOLD. The active site was assigned based on the volume occupied by the known ligand neopterin that was extracted from the complex structure. For docking processes in the RTA closed conformation, the receptor file was prepared from the apo RTA structure file (1RTC.pdb), and the ligand binding site was defined to cover the “secondary” pocket near Tyr80. The default setting was

used for all GOLD docking processes. In particular, 10 complete dockings were carried out for each compound with a total of 10,000 genetic algorithm operations exploring the conformational space. Each docking was followed by Simplex minimization in order to refine the ligand orientations to the nearest local minimum “Goldscore” was used as to rank the affinity of the ligand pose in the receptor. For user convenience, the score is adjusted to a positive value, that is, the higher the score, the better the predicted binding affinity. The cutoff score for binders might be around 50.

3.2.6 Docking with Surflex

Surflex was developed by Professor Jain and his colleagues at UCSF (Jain 2003; Jain 2007; Jain 2009). The program requires a “command” operation. All input molecules were protonated as expected at physiological pH. The active site was defined by the user as a “protomol” file; it could be generated in two ways, ligand-based and residue-based. In the preparation of the RTA open conformation receptor, the ligand-based method was used due to the availability of the neopterin ligand that was complexed with RTA in its open conformation, while the residue-based method was used to define the active site of RTA for its closed conformation. The docking process was launched by a line command that contained the defined active site, a file of compounds to be screened, and several parameters that set up the docking accuracy level, speed and information for the output file.

3.2.7 Docking with eHiTs

eHiTs automatically assigns and evaluates all of the possible protonation states for both the ligand and the receptor, so there is no need to add hydrogen atoms on the input molecules. The program uses an exhaustive search

algorithm, which breaks the potential ligand into distinct chemical groups, docks them independently, and rapidly enumerates mappings of interacting atoms between receptor and ligand. The binding energy of each pose is calculated and reported as a score. At the end of each run the ligand pose that produces the lowest score is saved and reported. Like Surflex, eHiTs uses command line to operate docking processes. The parameter `_clip` was used to set up active site of the receptor. The neopterin portion of the file 1BR5.pdb was used to map the ligand binding site. Docking with eHiTs was performed for open RTA conformation only.

3.2.8 Compound selection procedure

Two compounds were ultimately selected from the Sigma Aldrich library searches: (4-(3-(2-amino-4,6-dihydroxy-5-pyrimidinyl)propyl) benzoic acid (PBA) and 2-amino-1,4-dihydro-6-hydroxy-4-oxo-a-phenyl-pyrimidinepentanoic acid (PPA). These two compounds were within top 10 hits, based on ICM docking results for nearly 50,000 compounds. Also, the structure of the compound resembled that of other known RTA inhibitors in that key hydrogen bonds could be formed in the specificity pocket.

For the Chembridge Library selection process, the procedure was as follows: (1) All hits that were ranked within top 5% from ICM and GOLD; docking results were combined. (2) Initial hits were filtered through Lipinski's Rules (Lipinski, Lombardo et al. 2001) with an even more stringent solubility criterion. Specifically, the compounds should have less than 5 H-bond donors, less than 10 H-bond acceptors, less than 500 in molecular weight and less than 4 for the calculated Log P. (3) Similarity in structure was also considered for the filtering process. The compounds initially selected by the above rules were divided into

subfamilies according to structure resemblance, and structurally redundant compounds were reduced to a few representatives. By the above criteria, a total 306 compounds were selected and purchased from Chembridge dissolved in DMSO in 96 well plates.

3.2.9 Activity assay

The assay used to determine the inhibition efficacy of the selected compounds was an *in vitro* translation assay; protein synthesis was monitored by luminescence similar to the assay described in the previous section. However, for the Chembridge custom library of 306 “hit” compounds, the assay setup was modified due to the presence of DMSO (the compounds were dissolved in 100% DMSO at about 10mM upon arrival). 1 μ l of the above compound solution was mixed with 11 μ l BSA (1mg/ml), followed by a 30 min incubation at room temperature on the bench. Then, 4 μ l of 50 nM RTA (in HEPES pH 7.5) was added and incubated for another 30 min. At this point, the concentration of compound was about 625 μ M in 6.25% DMSO. Next, 1.5 μ l pre-incubated RTA/compound mixture was added into 7 μ l rabbit reticulocyte lysate, and the reaction was initiated by addition of 1.5 μ l of translation mix as described previously. The final concentration of compound in the assay was about 100 μ M in 1% DMSO. The protein concentration in the assay was about 2 nM, at which RTA was able to kill ~98% of control level protein synthesis.

3.2.10 Crystallization of RTA complexed with ligands

The compound (4-(3-(2-amino-4,6-dihydroxy-5-pyrimidinyl)propyl) benzoic acid (PBA) was dissolved in crystallization solution and soaked into monoclinic RTA crystals as described in the previous section, the final ligand concentration

was estimated at 5 mM. After a 5-day soak, the crystal was mounted for X-ray data collection.

The Chembridge ligands identified from the activity assay as hits were soaked into existing RTA crystals at different crystallization conditions that were obtained from previous crystal screenings using Hampton Research crystal screening kits (Crystal HT screen, Index HT screen, Peg/ion HT screen and Grid Salt HT screen).

3.2.11 X-ray Data Collection and Structure Determination

The instruments and software programs used were the same as previously described. The ChemBridge hits were soaked into crystal at three crystallization conditions. These reflected the solubility problems of the compounds and explored a range of pH and solvents. These conditions are (1) 0.1 M HEPES sodium pH 7.5, 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate monobasic (Hampton research Crystal HT screen C11) (2) 0.8 M Ammonium sulfate, 0.1 M Citric acid pH 4.0 (3) 1.4 M Sodium / Potassium phosphate pH 5.0 (4) 1.8 M Sodium / Potassium phosphate pH 5.0. The latter three are Hampton research Grid Salt HT screen conditions A1, F1, and F7 respectively.

3 RESULTS AND DISCUSSION

3.3.1 Docking performance of difference docking programs

eHiTs was the docking program we used in the early stage of VDS. In order to evaluate its docking performance, we used a small set of test compounds, comprised of known RTA inhibitors. The docking poses predicted for

all the known ligands resembled the binding poses in X-ray structures. However, the scores for the known RTA inhibitors were not well correlated with the actual binding affinity. Figure 16 showed the correlation of docking scores and the observed IC₅₀s.

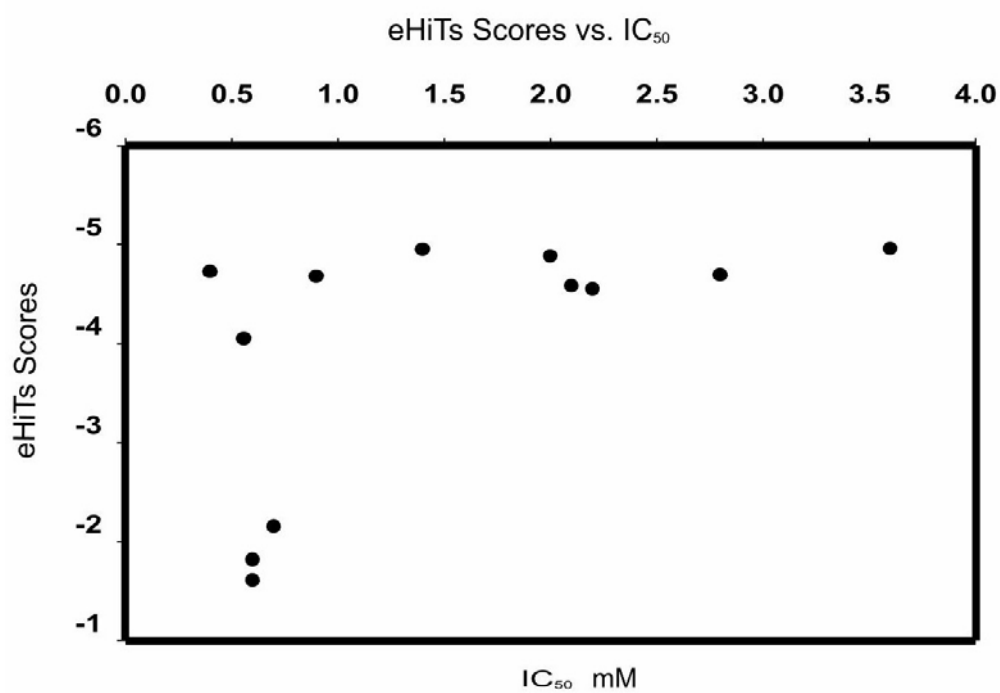


Figure 16. Correlation plot of eHiTs scores and IC₅₀s for known RTA inhibitors. 12 known RTA inhibitors were used to test the scoring performance of eHiTs. The X-axis is IC₅₀ in mM, the Y-axis is the docking score.

Since the scores are meant to approximate binding energies, the difference between scores should be very small for this set. More importantly there is no significant linear relationship between docking scores and observed IC_{50} s. Because of the poor performance of the ranking, this program couldn't distinguish real binders from decoys in the experiments we tried. We also observed that high molecular weight compounds seemed to have better scores than the compounds with low molecular weight. That is, adding atoms improved eHiTS scores regardless of stereochemical interactions. In addition, the early version of eHiTS tended to keep old docking results in memory, which interfered with subsequent runs. For these reasons, we did not continue using the program for RTA inhibitor searches.

Surflex was a fairly new docking program, developed at UCSF, when this project began. The program was known for its relative accuracy and docking speed (Welch, Ruppert et al. 1996). The algorithm is called Hammerhead, which is composed of three principal computational elements: an empirical scoring function that accurately predicts the affinities of a wide variety of known ligands given their crystallographic poses, a protein 'pocket finder' that automatically identifies binding regions, and a fragment-based alignment and conformational search procedure for small compounds. The high speed comes from the way that it filters through the poses. Specifically, some portion of a molecule that can make a specific interaction with the protein, above a threshold value, will be kept as a "head". For each possible head, the remaining molecular fragments are aligned one fragment at a time with difference poses. At each stage of fragment alignment, the best-scoring ligand arrangement will be retained. Therefore detailed computation was reserved only for the better potential binders. And

indeed, in our experiments, the docking poses for the known ligands were highly consisted with the crystallographic poses, and the docking speed was very impressive.

Unfortunately, Surflex lacks powerful post docking analysis tools. We used our own Perl scripts (written by Dr. Shuangluo Xia) to write out the best docking pose for each compound, and to extract the best poses from one huge output file. Another disadvantage of this docking program was that the parallel docking function was not originally available, and the program was inefficient for huge library docking tasks. For those reason, we did not do a thorough investigation of this program. Instead, we have mainly used two commercial docking programs for RTA inhibitor discovery, GOLD (Jones, Willett et al. 1995; Jones, Willett et al. 1997) and ICM (Abagyan and Totrov 2001). Direct and indirect comparisons have shown that GOLD consistently outperforms DOCK and FlexX in terms of average docking accuracy on a variety of systems (Verdonk, Cole et al. 2003). ICM, one of the more recently developed programs, has been reported to achieve a high degree of accuracy (Abagyan and Totrov 2001) and to provide the highest enrichment of the major programs (Warren et al., 2006).

The ICM program is based on a stochastic algorithm that relies on global optimization of the entire flexible ligand in the receptor field. That is, both the intramolecular ligand energy and the ligand–receptor interaction energy are optimized. The program combines several types of large-scale random moves, such as pseudo-Brownian moves, optimally biased moves of groups of torsions, and single torsion changes, with gradient local minimization and a history mechanism that both repels from the unwanted minima and promotes the discovery of new minima (Totrov and Abagyan 1997). The final poses are ranked

bases on the calculated interaction energy (vdW and electrostatic) minus the strain energy of the ligand conformation.

The GOLD program has several unique features, including that (1) it employs a bit-string ('chromosome') representation of conformations and possible hydrogen-bonding interactions between the ligand and receptor; (2) the ligand was set up as a fully flexible molecule, and the receptor was assigned partial flexibility for side-chains; and (3) special energy functions were derived, in part, from the analysis of conformation and non-bonded contacts observed in crystal structures of small molecules. One advantage of this evolutionary sampling technique is its ability to find solutions by exploring the full range of ligand conformational flexibility (Verdonk, Cole et al. 2003).

In order to evaluate how GOLD and ICM performed on our specific RTA project, we ran both programs on the Chembridge Fragment Library, containing about 4000 compounds. In addition, we added 11 known RTA inhibitors into the library as controls. We wanted to compare these two programs in two respects: the ability to identify the correctly docked pose for a known ligand and the ability of the scoring function to predict binding affinity. The docking results showed that both programs performed equally well at finding the correct binding poses for known ligands. About 65% of the ICM predictions for the control ligands bound within 2.0 Å RMSD (root mean square derivation) when compared with corresponding crystal structures. GOLD predictions were slightly worse than ICM, with about 60% of predicted poses within 2.0 Å RMSD from X-ray data. However, the ranking performances between the two programs differed substantially. There was little correlation between ICM score and GOLD scores; for the whole library the correlation coefficient was just -0.15, as shown in Figure 17. This

observation was, sadly, consistent with other reported docking program evaluation and analyses (I. Muegge 2001; Perez and Ortiz 2001; Taylor, Jewsbury et al. 2002).

Investigation of our results suggested that of the two sets of predictions, ICM was making the better choices. Specifically, ICM ranked 6 known ligands above the top 5% of all scores, 8 ligands above the top 10% and 10 ligands above top 20% in the ranking list. However, in GOLD, the rankings for the 11 controls were scattered from the highest rank of 0.03% to the lowest rank of 42.95%. Table 3 lists 11 control compounds and their corresponding rankings by ICM and GOLD. In addition, our results suggested that for GOLD, the predicted score versus affinity roughly tracks that for MW versus the predicted affinity. Figure 18 shows the correlation between molecular weights of the control compounds, with a correlation coefficient of 0.9. This observation has been pointed out by others for different docking programs (Roberts M. Stroud 2008). The cause is not clear but it suggests that configurational entropy of the ligands may be underestimated. ICM did not show such a correlation (correlation coefficient of 0.04).

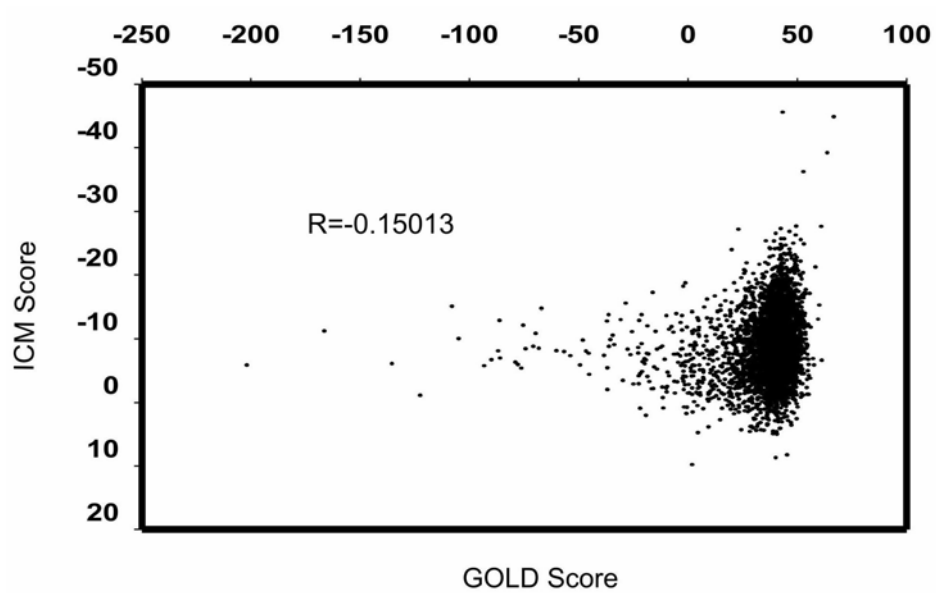


Figure 17. The correlation plot between GOLD and ICM scores for the Chembridge Fragment library.

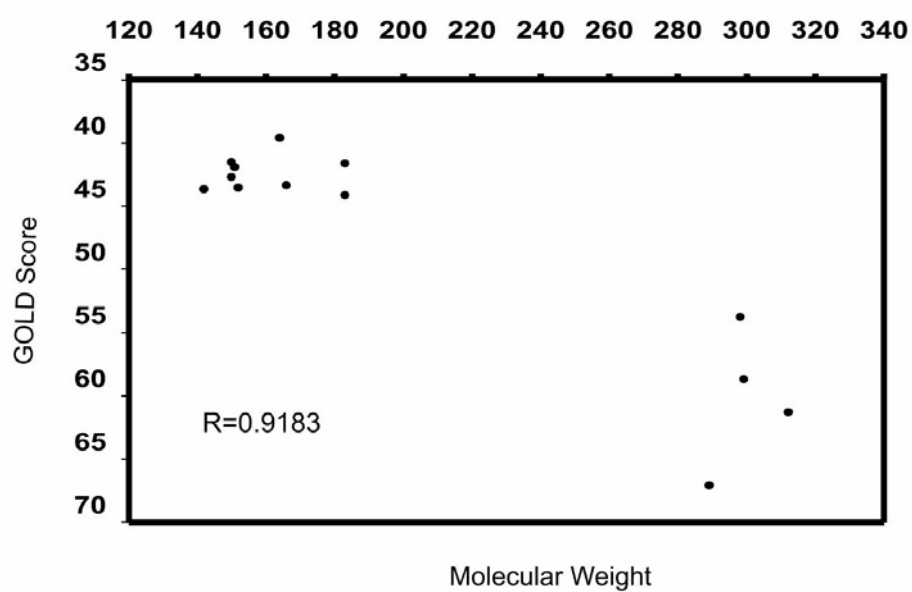


Figure 18. The correlation plot between GOLD scores and molecular weight of the control ligands used in the test.

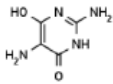
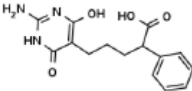
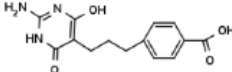
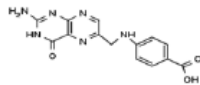
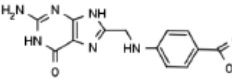
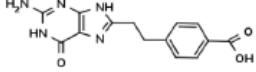
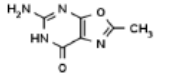
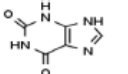
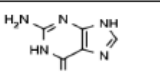
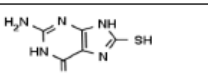
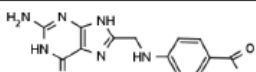
Ligand	ICM Rank (th)	GOLD Rank (th)	ICM top%	GOLD top%
	2	1152	0.05	29.75
	3	1	0.08	0.03
	4	2	0.10	0.05
	6	4	0.15	0.10
	64	7	1.61	0.18
	126	1225	3.17	31.64
	272	1663	6.85	42.95
	332	1183	8.36	30.55
	582	1589	14.65	41.04
	721	1048	18.15	27.07
	1137	37	28.63	0.96

Table 3. The predicted binding of known inhibitors used in the evaluation of GOLD and ICM docking. The ranking performance for these controls among 4000 compounds of the Chembridge fragment library are shown. The absolute rank and percentage from GOLD and ICM for each control compound are listed.

Based on these initial evaluation results, we chose ICM over GOLD as the main docking program for the RTA project. In addition to its apparently better docking performance, ICM has more powerful tools for post docking analysis. For example, it can compute the chemical properties of compounds, like logP, LogS, and druglikeness, which are very important parameters for users to make decision about which computer hits have potential drug lead value for further investigation. Also, the ICM program can categorize chemicals into families based on structural similarity, which is very useful for subsequent SAR (structure and activity relationship) analysis.

It must be pointed out, nonetheless, that ICM missed some active inhibitors from its top 15% hit list, which were picked up by GOLD. Therefore, in practice, we used both programs for each VDS task, and combined the top hits from each ranking list in order to catch as many potential hits as possible.

3.3.2 Virtual Screening of the Sigma-Aldrich Library

We used ICM to screen about 50,000 compounds from the Sigma-Aldrich library. This library contains several known RTA inhibitors, such as pterioic acid, guanine, and 7-deazaguanine (7DG); these compounds served as internal controls and benchmarks for hit docking and scoring. The screening with ICM showed that the known inhibitors pterioic acid, guanine, and 7DG ranked 40th, 8th, and 6th respectively in the scoring list of 50,000 (Table 4). That is, the known RTA inhibitors all ranked in the top 0.08% of the screen. However, GOLD the rankings for these known RTA inhibitors were very low. As with the Chembridge screen, it appeared that ICM was a more useful screening program for RTA than was GOLD.

An examination of the top 10 hits from the virtual screening by ICM showed they fell into three structural classes, one pterin, four guanines and five pyrimidines. Guanine and pterin derivatives are well-known RTA inhibitors binding at the active site (Miller, Ravikumar et al. 2002). However, only one pyrimidine compound, DDP, was previously reported as an RTA inhibitor, and the X-ray structure showed that it binds RTA in a Tyr80-closed conformation (Miller, Ravikumar et al. 2002).

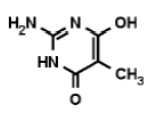
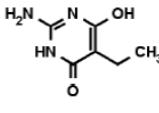
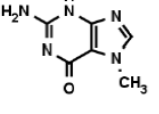
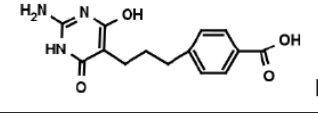
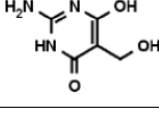
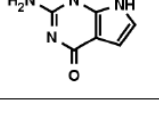
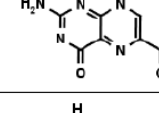
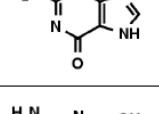
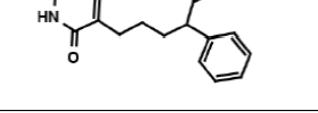
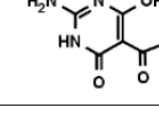
	Structure	ICM_rank Top%	GOLD_rank Top%
1		0.0020	61.65
2		0.0041	53.96
3	 7MG	0.0061	30.19
4	 PBA	0.0082	1.16
5		0.0102	31.32
6	 7DG	0.0122	51.58
7		0.0143	8.80
8	 Guanine	0.0163	67.03
9	 PPA	0.0184	1.19
10		0.0204	51.42

Table 4. Top 10 hits from Sigma-Aldrich library by ICM. The compound structures and their corresponding rankings in ICM and GOLD are listed.

However, the ICM virtual screening was performed with RTA in the open conformation, predicting that these compounds interact favorably with groups within the RTA specificity pocket. Another interesting observation was that two pyrimidine compounds, PBA and PPA ranked among the top hits from ICM docking were also ranked high by GOLD, 1.16% and 1.19% respectively. All these observations suggested that there were some important and useful properties about the pyrimidine moiety with respect to binding. To test this, we purchased two of the pyrimidines that had pendants reminiscent of pteronic acid (PTA). These compounds, (4-(3-(2-amino-4,6-dihydroxy-5-pyrimidinyl)propyl) benzoic acid (PBA) and 2-amino-1,4-dihydro-6-hydroxy-4-oxo-a-phenyl-pyrimidinepentanoic acid (PPA) were selected to test in the activity assay.

3.3.2.1 Activity and structure of PBA

Both PBA and PPA, introduced above, inhibited RTA with IC_{50} values of 0.27 mM (Fig.19) and >1mM respectively (Table 6). (PPA itself showed inhibitory effects on protein synthesis at concentrations above 1mM). In order to understand the binding mode, we soaked the ligands into pre-existing monoclinic RTA crystals. Crystallographic data for the RTA/PBA structure is shown in Table 5. Figure 20 shows the electron density map for the PBA ligand bound into the RTA active site. It clearly showed the binding mode described as follows: (1) the aromatic ring of PBA appears to form an aromatic stack with the side chain of Tyr80, (2) the exocyclic amino group on C₂ donates hydrogens to the carbonyl oxygens of Val81 and Gly121, and (3) one ring nitrogen, N₁, is unprotonated and can accept a hydrogen bond from the amido group of Val81, while the other ring nitrogen, N₃, is protonated and donates a hydrogen bond to the carbonyl oxygen of Gly121. A unique aspect of this inhibitor is the acidity of the alcohol group at

C₆. The pKa is estimated to be 6.5 (MarvinSketch, ChemAxon Ltd) and the group is ionized at physiologic pH, allowing the alcoxide ion to pair with the side chain of Arg180, shown in Figure 21.

The ICM predicted binding pose, shown in Figure 22, is well matched with the position observed in the crystal structure. The pyrimidine rings are essentially identical, although the benzoic acid tail differs between the two. However, the electron density of this part of the molecule is very weak, indicating that it is in motion, so it is possible that the benzoic acid group can be positioned as predicted at least transiently. Figure 23 shows a superposition of PBA and PTA from their respective crystal structures; it is clear they occupy the same space in the active site and make many of the same specific hydrogen bond interactions in the RTA specificity pocket.

PTA, built on the larger pterin platform, is a more rigid inhibitor than PBA; there are only three rotatable bonds in PTA and four in PBA. This means that PBA binding may be slightly disfavored entropically. Increasing the rigidity of future inhibitors based on PBA will be a major design goal.

The pyrimidine platform, however, has some apparent advantages for inhibitor design compared with the pterin based compounds. The major one is solubility. The platform for PTA is 2-amino-1,4-dihydropteridin-4-one, and that for PBA is 2-amino-6-hydroxy-3,4-dihydropyrimidin-4-one. Analysis by a number of computational programs suggests that the latter is at least ten times more soluble than the former in aqueous solutions. For example, the AlogS for the pterin derivative is -2.0 (1.6 g/l) while that for the pyrimidine base compound is -1.0 (11.6 g/l). This trend carries over to the inhibitors as well, where PTA has a value of -3.5 (0.1 g/l) while PBA has a value of -2.1 (2.3 g/l).

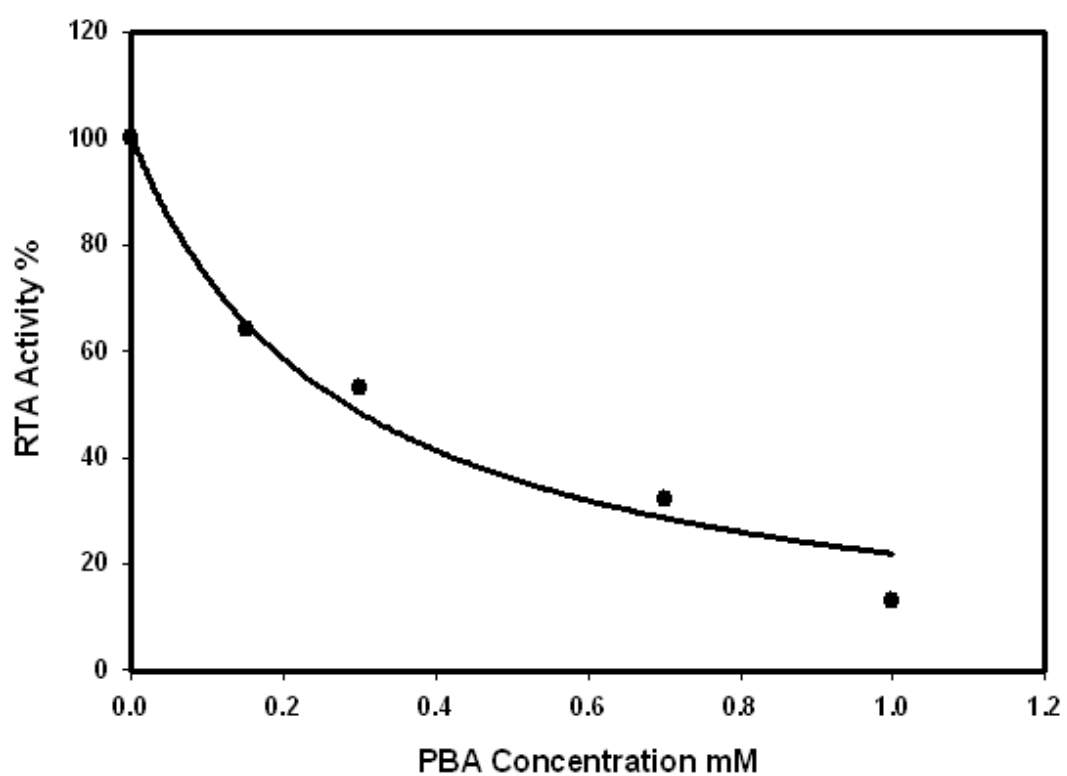


Figure 19. Inhibition of RTA by PBA. The activity of RTA is plotted as a function of PBA concentration. The data were fitted to a hyperbola, revealing an IC_{50} of 0.27 mM. About 0.2 nM RTA in 20 mM HEPES (pH7.5) was used in the assay; the PBA concentrations used were 0.15, 0.3, 0.75, and 1 mM.

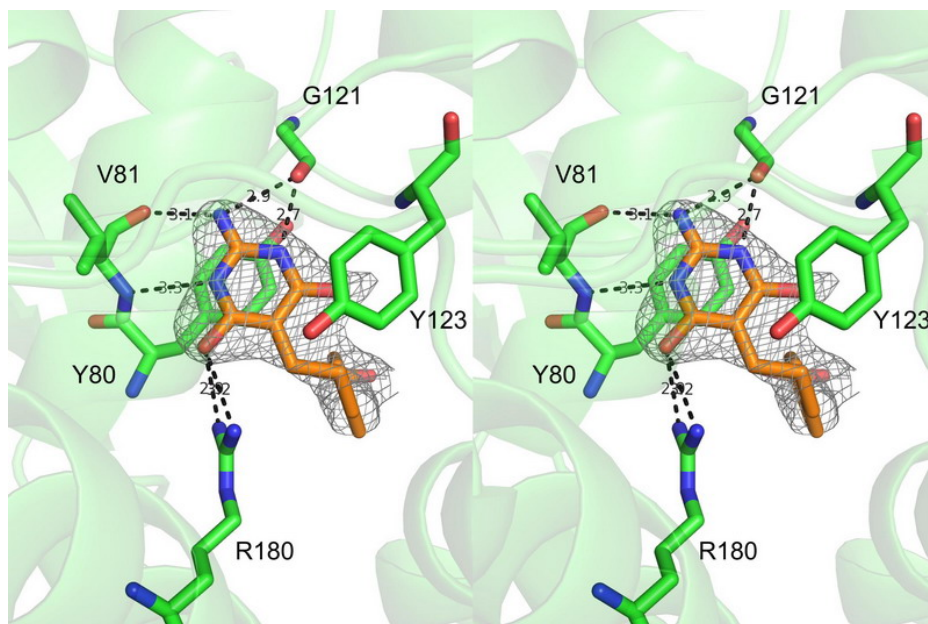


Figure 20. Stereo view of the omit electron density map for PBA in RTA•PBA complex. PBA is in orange, protein in green. Hydrogen bonds are indicated by dotted lines. Active site residues are labeled in one letter code.

Statistics	RTA-PBA
a (Å)	42.6
b (Å)	68.0
c (Å)	49.7
α (deg)	90.00
β (deg)	113.0
γ (deg)	90.00
Space group	$P2_1$
Resolution (Å)(last shell)	27.5-2.5
No. obsd	31332
R_{merge} (last shell) (%)	4 (11.3)
I/\bar{O}_1 (last shell) (%)	30.1 (8.2)
Completeness (last shell) (%)	99.8(99.8)
Redundancy	3.4
No. reflections	8469
R_{work} (%)	19.5
R_{free} (%)	26.6
Rms deviation from ideality (Å,deg)	0.010/1.237
Estimated overall coordinate error (Å)	0.337
Correlation coefficient F_o-F_c	0.933
Correlation coefficient $F_o-F_{c\text{free}}$	0.896

Table 5. X-ray data. Data collection and refinement statistics for RTA/PBA complex.

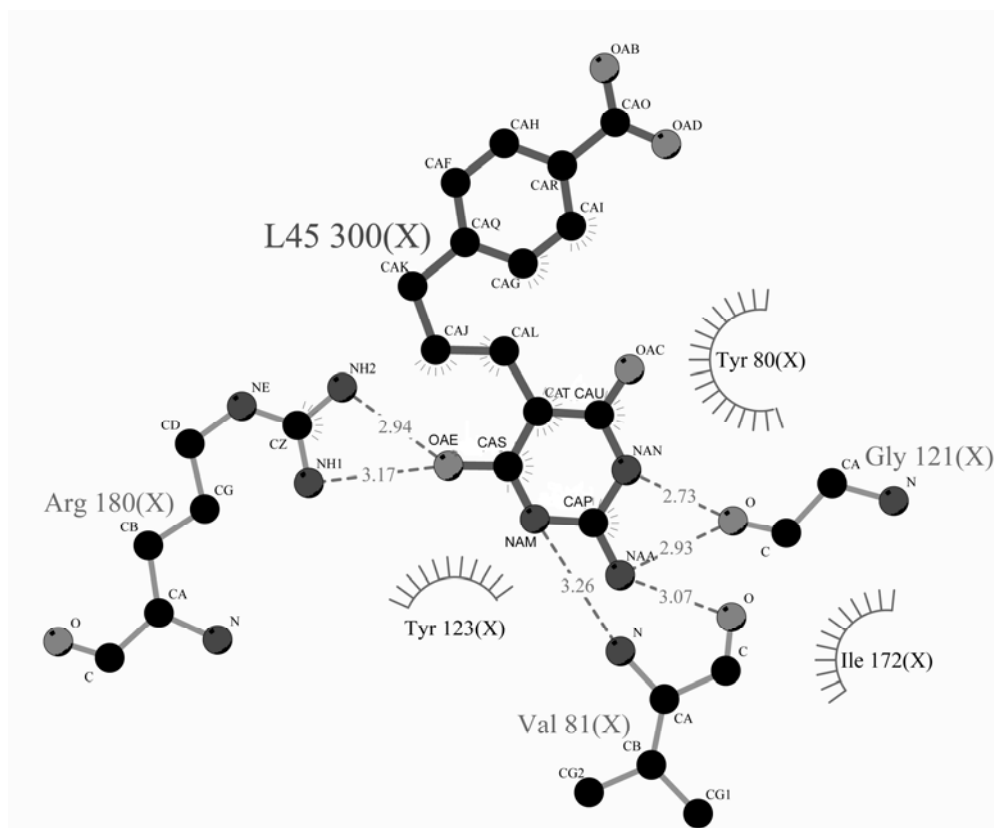


Figure 21. Interactions between PBA and RTA. The hydroxyl group at the position 4 is negatively charged at the assay and crystallization pH conditions. Hydrogen bonds are shown as dashed lines with lengths indicated. Hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. This figure was made with LIGPLOT (Wallace, Laskowski et al. 1995)

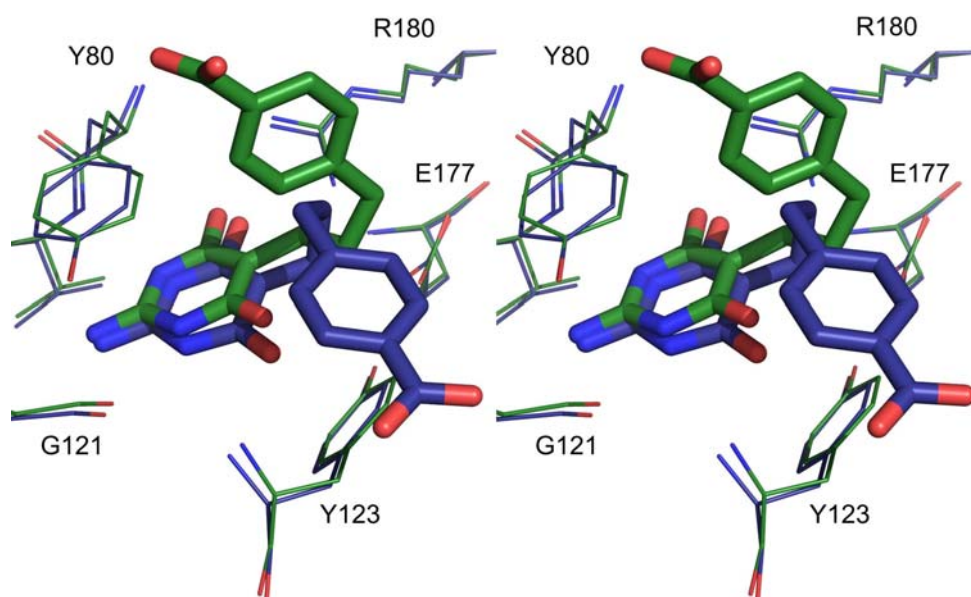


Figure 22. Superposition of the ICM predicted PBA binding pose with the X-ray structure of RTA•PBA. The ICM prediction is in dark blue, and PBA crystal structure is in green. Active site residues are labeled with the one letter code.

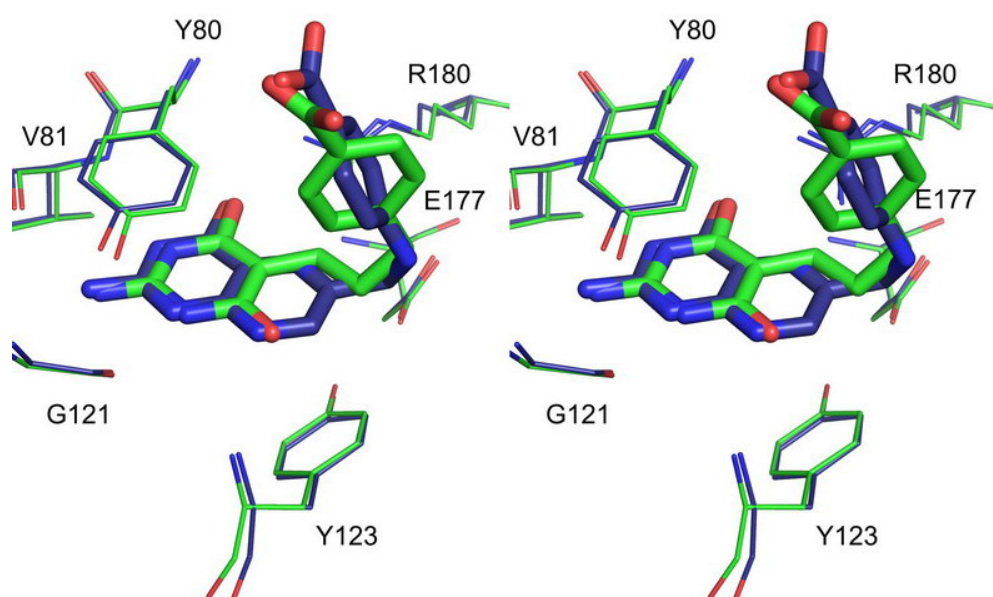


Figure 23. Superposition of the active sites of crystal structure for RTA•PBA and RTA•PTA. RTA•PBA is in green, and RTA•PTA (1RB6.pdb) is in dark blue. Active site residues are labeled with the one letter code.

3.3.2.2 Structure and activity relationship between PBA derivatives

Given our success with PBA, we tested 9 additional pyrimidine derivatives, which resulted in 3 more RTA inhibitors, all with weaker inhibitory effects than PBA. The 10 tested compounds (shown in Table 6) have a common 2-amino-dihydroxy pyrimidine ring and different branching groups at the carbon 5 position; these branching groups exhibit different pendant lengths, polarity, and size. Compounds 1 (PBA) and 2 (PPA), are RTA inhibitors, possessing a bulky aromatic group and a carboxylate group at the end of the tail. The virtual docking results indicated that the aromatic ring may contribute hydrophobic interactions that facilitate binding, and that the carboxylate may form polar contacts with RTA, or may simply increase the compounds solubility and effective concentration in the assay. There is chemical similarity between compounds 5, 6 and 10. However only compounds 6 and 10 showed measurable inhibition of RTA, suggesting that a three or four-methylene linker is tolerable between the pyrimidine and other functional groups on the inhibitor. Compound 8 did not inhibit RTA activity, indicating the 2-amino pyrimidine ring itself is not enough to form strong interactions with RTA.

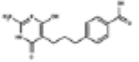
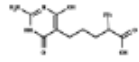
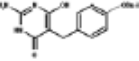
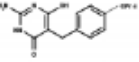
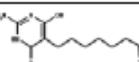
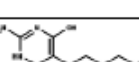

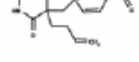
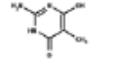
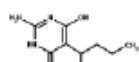
	2D Structure	Name	IC ₅₀
1		4-[3-(2-amino-1,4-dihydro-6-hydroxy-4-oxo-5-pyrimidinyl) propyl]-benzoic acid	0.27mM
2		2-amino-1,4-dihydro-6-hydroxy-4-oxo-a-phenyl-5-Pyrimidinepentanoic acid	>1mM
3		2-amino-6-hydroxy-5-[[4-(2-methylpropoxy)phenyl]methyl]-4(1H)-Pyrimidinone	No Inhibition
4		2-amino-6-hydroxy-5-[[3-methyl-4-(1-methylethoxy)phenyl]methyl]-4(1H)-Pyrimidinone	No Inhibition
5		2-amino-1,4-dihydro-6-hydroxy-4-oxo-5-Pyrimidineheptanoic acid	No Inhibition
6		2-amino-1,4-dihydro-6-hydroxy-4-oxo-5-Pyrimidinepentanoic acid	>3mM
7		4-((5-ALLYL-2-imino-4,6-dioxohexahydro-5-pyrimidinyl) methyl)benzoic acid	No
8		2-amino-6-hydroxy-5-methyl-4(1H)-Pyrimidinone	No Inhibition
9		2-(2-amino-4,6-dihydroxy-5-pyrimidinyl) pentanoic acid	No Inhibition
10		((5-(2-amino-4,6-dihydroxy-5-pyrimidinyl)pentanoyl)amino)acetic acid	2mM

Table 6. PBA derivatives. There are 10 PBA derivatives tested in this work. Their structures, names and activities are listed here. Compound 1 and 2 are PBA and PPA respectively.

Based on these activity information for pyrimidine derivatives, we went back to evaluate the performance of the docking programs. On the correlation graph of ICM score vs. IC_{50} (Fig. 24), all four real hits are scored lower than -32, which is the default cutoff for binders. However, for nonbinders, the score could vary widely, indicating that the scoring function is still far from precise in discriminating decoys from the real binders. In the plot an IC_{50} of 100 mM has been arbitrarily assigned to the nonbinders in order to complete the plot. In addition, the correlation of rankings between ICM and GOLD is poor (Fig. 25). This observation is not just for the 10 inhibitor compounds, but for all the large libraries that we have tested. However, we also observed that although the ranking between programs didn't correlate well for the whole library, the docking poses and rankings for real ligands were correlated, at least for the small tests we have done. For example, the four hits listed here are all ranked highly by both programs. Figure 25 showed the correlation of ICM scores and GOLD scores for the four real hits (in red dots). Figure 26 showed that the docking poses for these four hits were consistent between ICM and GOLD. Therefore, the strategy for virtual screening would be: run multiple docking programs, combine top 5% compounds from each program, keep the common ones, and keep those with reasonable docking poses, that is, those that made sensible polar interactions.

Unfortunately, among the four inhibitors derived from the pyrimidine platform, only PBA has crystallized with RTA as a complex. It could be that the binding affinity of the other three ligands is so weak that they could not be fixed in the complex form under the crystallization conditions.

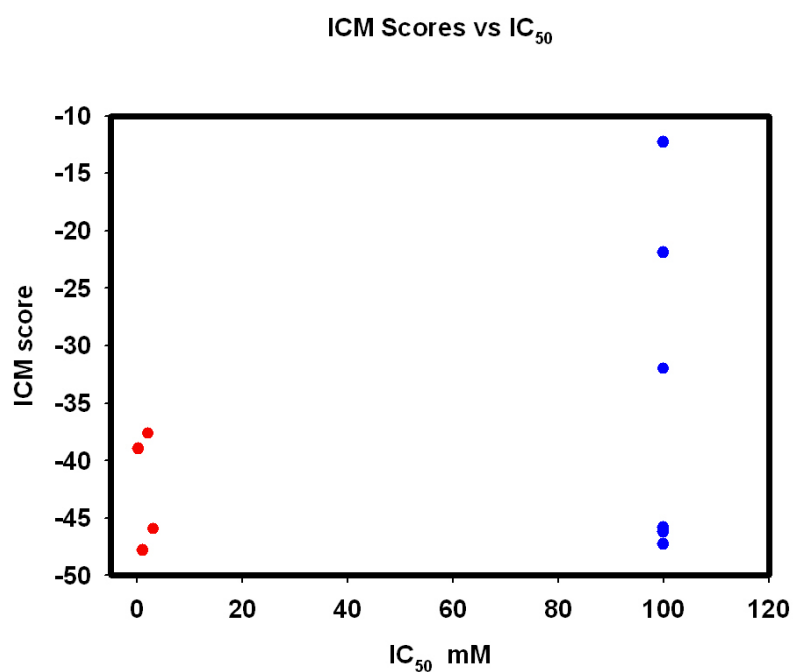


Figure 24. The correlation between ICM score and IC₅₀. An IC₅₀ of 100 mM has been arbitrarily assigned to nonbinders in order to make the plot. Four real binders are shown in red.

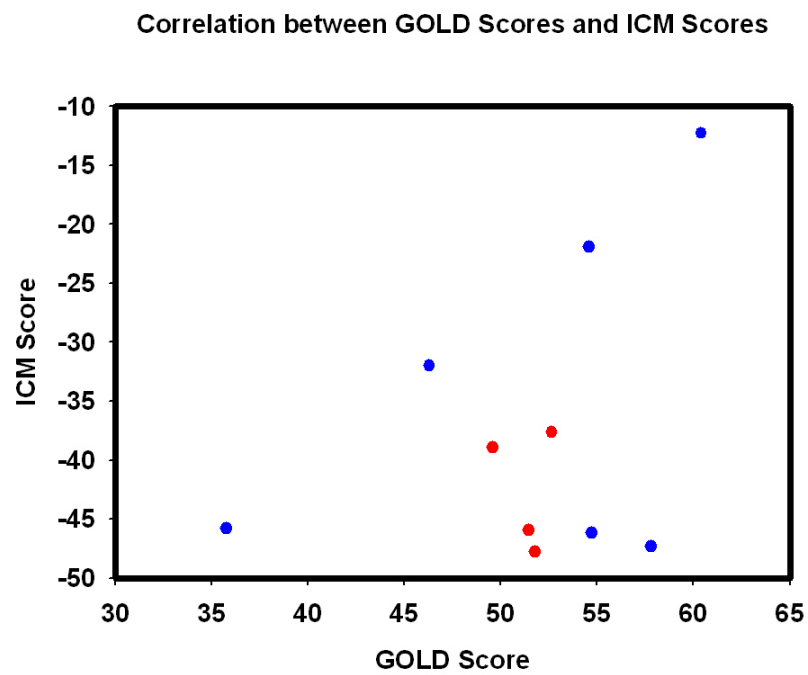


Figure 25. The Correlation of GOLD score and ICM score. The correlation coefficient is 0.2738. Four real hits are shown in red.

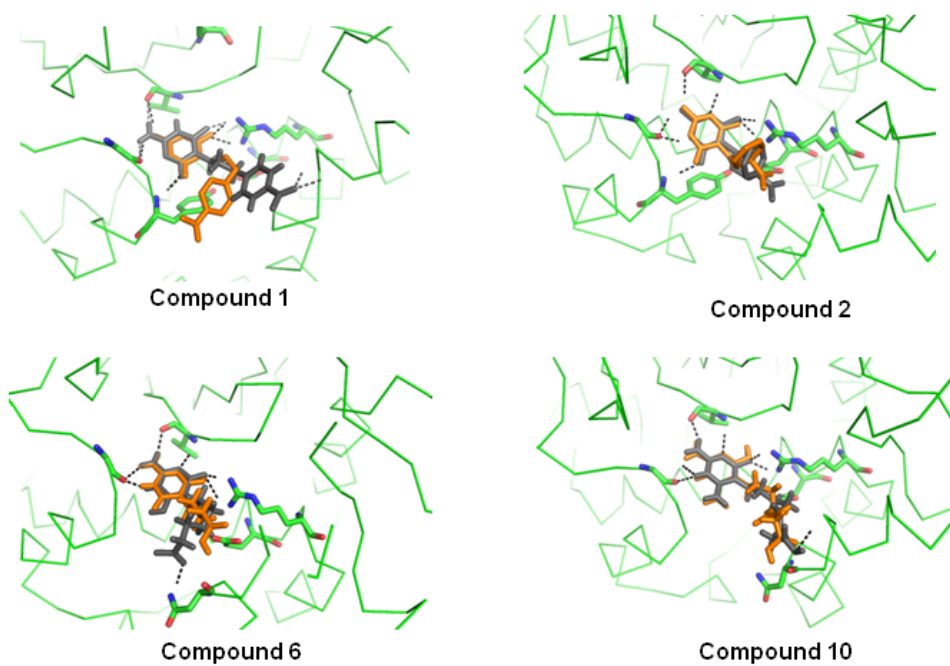


Figure 26. Docking poses of RTA inhibitors derived from PBA. The docking poses predicted by GOLD are shown in grey, and those predicted by ICM are shown in orange.

3.3.2.3 Optimization of pyrimidine inhibitors

RTA crystal structures showed that there is a second pocket adjacent to the adenine binding specificity pocket, which might accommodate a second chemical group (Yan, Day et al. 1998) This pocket is close enough to the specificity pocket, so that pendants attached to the specificity group might bind there. Two strategies are ongoing to find suitable pendants for the second binding site. One is to elongate PBA at its benzoic acid end with commercially available amino acids by a simple peptide coupling reaction, shown in Figure 27. Another way is to use a fragment-based drug discovery strategy to find the fragments that can bind in the second pocket and then link the pendant to the pyrimidine platform by an appropriate chemical reaction.

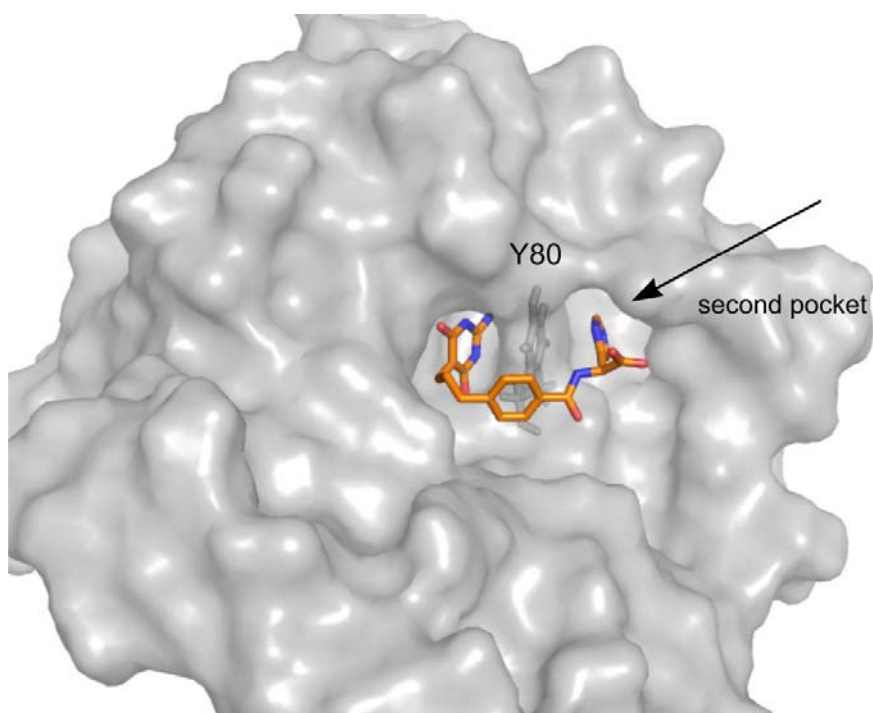


Figure 27. The model of histidine coupled PBA bound at two RTA pockets. The second pocket is indicated by the arrow. The Tyr80 side chain is shown in grey sticks. PBA coupled with histidine is in orange. The 2-amino-dihydroxy pyrimidine head of the ligand binds at the specific pocket and the histidine side chain sits in the secondary pocket.

We collaborated with Dr. Anslyn in the Department of Chemistry and Biochemistry to synthesize the amino acid coupled pyrimidine derivatives. We made the models for PBA coupled with every individual amino acid types and chirality, and docked them into RTA using ICM. Table 7 lists the docking results, which suggested that a number of putative derivatives can reach the second pocket. Since we have seen a small ring structure (DDP) bound in the second pocket in X-ray structure previously (Miller, Ravikumar et al. 2002), it might be possible that aromatic amino acids would go to the second site preferentially. Figure 27 illustrates how the histidine coupled PBA would block both the specific pocket and the second binding site nearby. In particular, the side chain of histidine sits the second site by stacking with Tyr80 side chain. Although the carboxylate groups of the putative derivatives would be unfavorable for cell permeability, that problem could be overcome in drug development by changing this group to other bioisosteres, like tetrazoles. The chemistry pathways for those changes are well-established (Biot, Bauer et al. 2004).

The second site is a relatively large pocket that is lined at the surface with polar side chains but with potential hydrophobic contacts deep inside. Therefore, the binding of small ligands in this pocket might not be specific. Fragment-based searching method may be a good way to identify potential ligands for this second pocket. It could provide many small functional scaffolds with different binding preferences in the second pocket, which may help map the whole area. It is known that considerable gains in ligand potency can be achieved through linking together fragments that bind independently in different regions of the protein

(Jencks 1981; Murray and Verdonk 2002; Edward R. Zartler 2008). The work that has been done using this method is discussed in the following section.

NAME	Score	MolLogP	DrugLiker	MolLogS
45_gln_R	-44.78	-0.237	0.351	-3.615
45_tyr_S	-44.59	1.583	0.808	-4.542
45_thr_S_R	-42.91	-0.174	0.642	-2.578
45_phe_R	-40.87	1.843	0.558	-4.685
45_trp_S	-38.57	2.060	-0.070	-6.103
45_phe_S	-38.51	1.700	0.904	-4.632
45_tyr_R	-38.45	1.583	0.808	-4.542
45_glu_S	-37.65	0.362	0.300	-3.166
45_his_S	-37.36	-0.062	0.745	-3.907
45_his_R	-36.25	-0.062	0.745	-3.907
45_ser_S	-37.36	-0.432	0.341	-2.996
45_ser_R	-35.25	-0.432	0.341	-2.996
45_cys_R	-35.04	0.733	0.506	-3.834
45_arg_R	-34.84	0.020	0.112	-3.485
45_ala_R	-34.39	0.633	0.486	-3.146
45_met_S	-34.36	1.212	0.394	-4.039
45_asp_R	-34.16	-0.120	0.186	-2.688
45_trp_R	-33.88	2.060	-0.070	-6.103
45_asp_S	-33.61	-0.120	0.186	-2.688
45_leu_S	-32.92	2.012	0.517	-4.335
45_lys_S	-31.77	1.707	0.184	-3.570
45_met_R	-30.73	1.212	0.394	-4.039
45_asn_S	-30.03	-0.719	0.275	-3.059
45_gln_S	-29.92	-0.237	0.351	-3.615
45_cys_S	-29.20	0.733	0.506	-3.834
45_val_R	-28.94	1.463	0.192	-4.147
45_ala_R	-28.76	0.633	0.486	-3.146
45_asn_R	-28.22	-0.719	0.275	-3.059
45_lys_R	-27.91	1.707	0.184	-3.570
45_ile_S_R	-27.79	2.040	0.294	-4.516
45_ile_S_S	-26.97	2.040	0.294	-4.516
45_val_S	-26.66	1.463	0.192	-4.147
45_arg_S	-26.09	0.020	0.112	-3.485
45_glu_R	-26.08	0.362	0.300	-3.166
45_ile_R_R	-26.00	2.040	0.294	-4.516
45_leu_R	-25.50	2.012	0.517	-4.335
45_ser_S	-24.75	-0.432	0.341	-2.996
45_ile_R_S	-23.71	2.040	0.294	-4.516
45_ser_R	-15.84	-0.432	0.341	-2.996
45_thr_R_R	-9.20	-0.031	0.297	-2.631

Table 7. ICM docking results for amino acid coupled PBA. 'R' and 'S' represent chirality of amino acids. Solubility and drug likeness computed by ICM are listed.

3.3.3 Virtual Screening of the Chembridge library

As described in the methods section, 306 compounds were selected and purchased from Chembridge based on their rankings by ICM and GOLD. The dockings were performed for both RTA open and closed conformations. We wanted to test representatives of both binding classes in the hope of finding novel scaffolds for both the specific pocket binding and the second site binding. The in vitro translation assay was done in the presence of 1mg/ml BSA to prevent nonspecific binding caused by aggregates that certain small molecules would form at high concentration (Seidler, McGovern et al. 2003; Feng, Shelat et al. 2005). Five compounds showed measurable RTA inhibition activity and four of them showed inhibition comparable with previous RTA inhibitors, that is at a high μM levels. These five compounds were also subject to activity tests for SlitA inhibition; two compounds inhibited SlitA with high μM level IC_{50} values. Table 8 lists the structures of these hits and their IC_{50} values. Crystallography showed that three hits are able to be crystallized with RTA as complexes; their structures are currently undergoing further refinement.

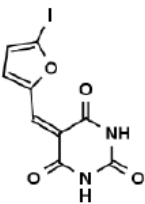
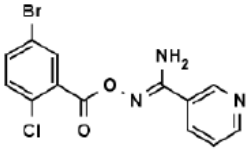
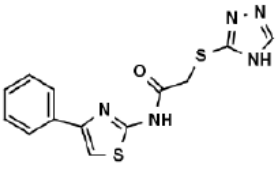
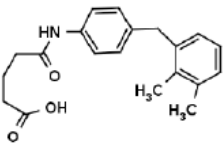
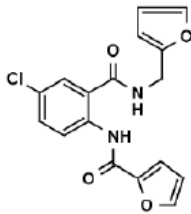
Structure	Name	IC ₅₀ μM (<i>in vitro</i>)	Resolution Å
	5-[(5-iodo-2-furyl)methylene]-2,4,6 (1H,3H,5H)-pyrimidinetrione 5225540	125 (RTA) toxic > 250	1.5
	N'-[(5-bromo-2-chlorobenzoyl)oxy]]-3-pyridinecarboximidamide 7780474	>2000 (RTA) 1000 (SIt)	1.7
	N-(4-phenyl-1,3-thiazol-2-yl)-2-(4H- 1,2,4-triazol-3-ylthio)acetamide 5978818	500 (RTA)	N/A
	4-{[4-(2,3-dimethylphenoxy)pheny l]amino}-4-oxobutanoic acid 5282931	200 (RTA) toxic>125 uM	N/A
	N-(4-chloro-2-[[[(2-furylmethyl)ami no]carbonyl]phenyl]-2-furamide 7543758	250 (RTA) 400 (SIt)	1.4

Table 8. Hits identified from Chembridge library by VDS. Their structures, names, activity (*in vitro*) and X-ray data resolutions are listed in the table.

Compound 5-[(5-iodo-2-furyl)methylene]-2,4,6(1H,3H,5H)-pyrimidinetrione (Chembridge ID: 5225540) showed RTA inhibition both *in vitro* and *in vivo* (cell-based assay, see details in the following chapter). Figure 28a. shows the dose response curve of the compound in an *in vitro* translation assay. The assay was done by varying the ligand concentration with a range of 30 μ M to 500 μ M with or without RTA. The assay results suggested that the ligand becomes highly toxic to the protein synthesis system at high concentration (above 250 μ M). Since there were two phenomena occurring in the assay, inhibition of protein synthesis by the compound and by RTA, it was difficult to calculate the exact IC₅₀. However, the compound did show increasing RTA inhibition in a dose dependent manner with a maximum inhibition about 40% at 125 μ M. Considering the inhibition effect on the protein synthesis caused by the ligand at this concentration the actual inhibition of RTA may be higher than shown in the plot. Figure 28b shows the dose curve for the ligand in a cell-based assay (24h and 48h incubation time) carried out by our colleagues at the Wadsworth Center in New York. The results are consistent with our *in vitro* assay results in terms of dual effect on the assay system and on RTA itself. Specifically, after 24 hours incubation of the cells with 90 μ M ligand only about 90% of cells remained viable. This is about the same number of cells that remained alive in the presence of both the compound and RTA, while RTA alone killed 70% of the cells. However, it seemed that cell toxicity of the compound increased with the increase of incubation time, as shown in figure 28 c. With only 25 μ M compound, the cell viability remained the same (about 60%) with or without RTA, which was consistent with the observation of RTA inhibition in 24h long experiment. The IC₅₀ from the cell-based assay was more than 10 times lower than that of *in vitro* assay, suggesting

that the compound was somehow modified to a more effective format during metabolism, or that the IC50 values, unlike K_i values, are a function of the assay format, especially incubation time.

5225540 Dose Curve

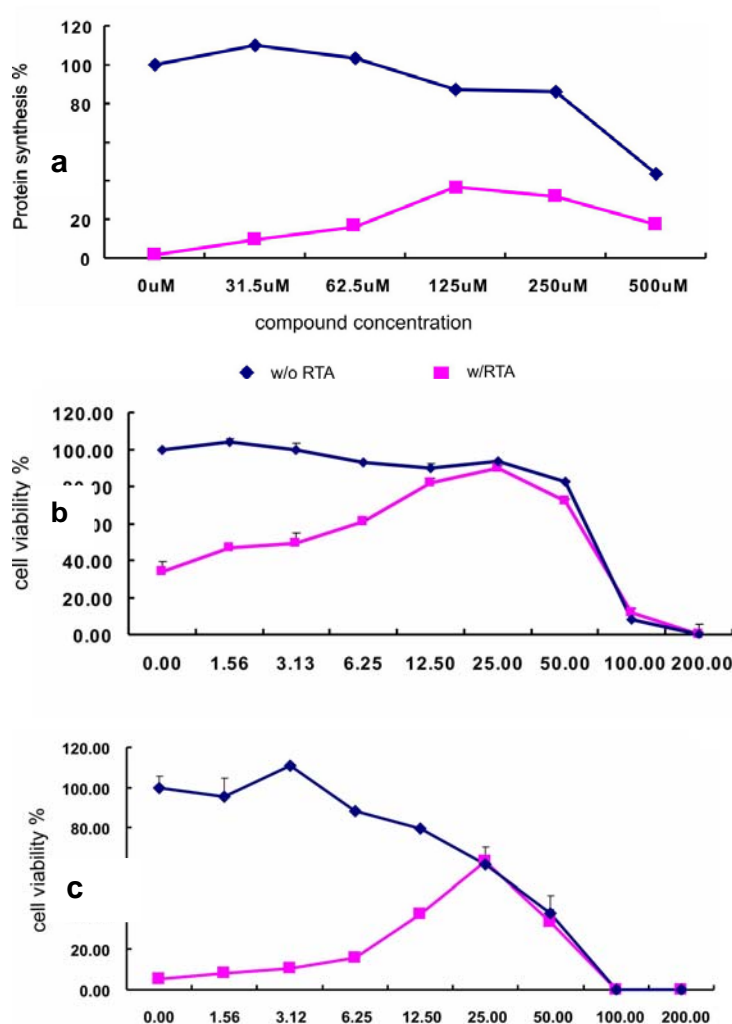


Figure 28. Compound 5225540 dose response *in vitro* and in Vero cells. a. Protein synthesis altered by compound itself, and by the compound and RTA. b and c show the cell viability changes with addition of the compound only, and the compound with RTA together. Blue lines depict the data for the compound only; pink lines show the effect when both the compound and RTA are present.

Our preliminary crystallization results appeared to confirm that compound 5225540 binds to the active site of RTA and acts as a novel inhibitor. Figure 29 shows the X-ray structure of RTA soaked with the compound. The ligand was soaked into RTA monoclinic crystals as described in the methods section. The electron density for the Tyr80, omitted in the initial refinement model, was not found in the native closed conformation, but was displaced by the inhibitor to the open position seen with other specificity site ligands. In addition, there was electron density in the difference map within the specificity site, indicating the presence of the inhibitor. However, the density for the ligand was broken and discontinuous, which might suggest binding was not strong enough to occupy the active site fully, or that the binding could occur with multiple modes. Based on the current map, it is a real challenge to build the ligand with a single convincing binding mode, therefore the crystallization condition needs to be further refined.

In addition to compound 5225540, compounds 7543758 and 5978818 inhibit RTA in our *in vitro* assay with IC_{50} values of about 250 μ M and 500 μ M respectively. Compound 7780474 showed very weak inhibition against RTA with IC_{50} greater than 2 mM, but showed slightly stronger inhibition against SlrA with an IC_{50} of 1mM. Compound 5282931 showed RTA inhibition with IC_{50} about 200 μ M, but it also strongly inhibited protein synthesis in the assay, as least when the concentration was greater than 125 μ M. Neither compound 7543758 nor 5978818 inhibited ricin in the cell-based assay, probably because of their poor cell permeability and cell toxicity. Compound 7780474 has not been tested in cells yet. All these three inhibitors were subject to crystallization by soaking into RTA crystals at different crystallization conditions.

We have collected two data sets for RTA crystals soaked with compound 7543758. One was crystallized in monoclinic conditions, and another one was in the tetragonal crystallization condition (0.1 M HEPES sodium pH 7.5, 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate monobasic). In both data sets, the side chain of Tyr80 was displaced from its native conformation, and some difference density showed up in the specificity pocket. Figure 30 shows the omit difference map for the complex of structure of RTA•7543758. Although the electron density for the ligand was very weak, the density for Tyr80 side chain clearly suggests the protein active site was open to accommodate the ligand. The electron density for the complex RTA•7780474 suggests a similar situation for this ligand, shown in Figure 31. We did not see any ligand density for compound 5978818 after refinement of two data sets, although the inhibition of this compound was stronger than that of 7780474.

The failure of a known ligand to bind with its receptor protein in the crystalline state is not uncommon. There are many possible reasons for this frustrating behavior. Sometimes, the ligand that shows up in an activity assay may not be a real ligand but a false positive; as such it of course will never be seen in complex crystals. In order to rule out this possibility, a secondary assay, usually a biophysical binding assay, is employed to further confirm the binding. We tried several methods, besides the current translation activity assay, to measure physical interaction of RTA and the inhibitor candidates. Preliminary data indicate that isothermal titration calorimetry, ITC, might be a good method, and it is discussed in the following chapter. Failure to generate a small molecule-protein complex are frequently due to major differences between the assay conditions and those necessary to generate high concentrations of

homogeneous complex for crystallization trials. Success may also be influenced by other operational details, such as how the ligand is mixed with protein solution, how the ligand is dissolved, how much ligand is added and how long the soaking process lasts. For weaker complexes, as in our case, there would be a substantial proportion of free protein within the crystallization drop and it is possible that the unbound protein crystallize first, and precludes addition of complexes with subtly different structures (Chung 2007).

In our case, we have seen ligand density in complexes for three hits discussed above, arising from virtual screening; this greatly strengthens the notion that they are real ligands for RTA. Next, we need to analyze factors that can improve ligand binding in the crystallization conditions. There are several things worth trying. For instance, we can use different solvents to increase compound solubility in crystallization solutions.. We can also screen co-crystallization conditions that would be more favorable for complex formation. For poorly soluble compounds, the solution could be initially made at low concentration, and then concentrated, together with protein, which may keep the ligand in solution throughout the concentration process.

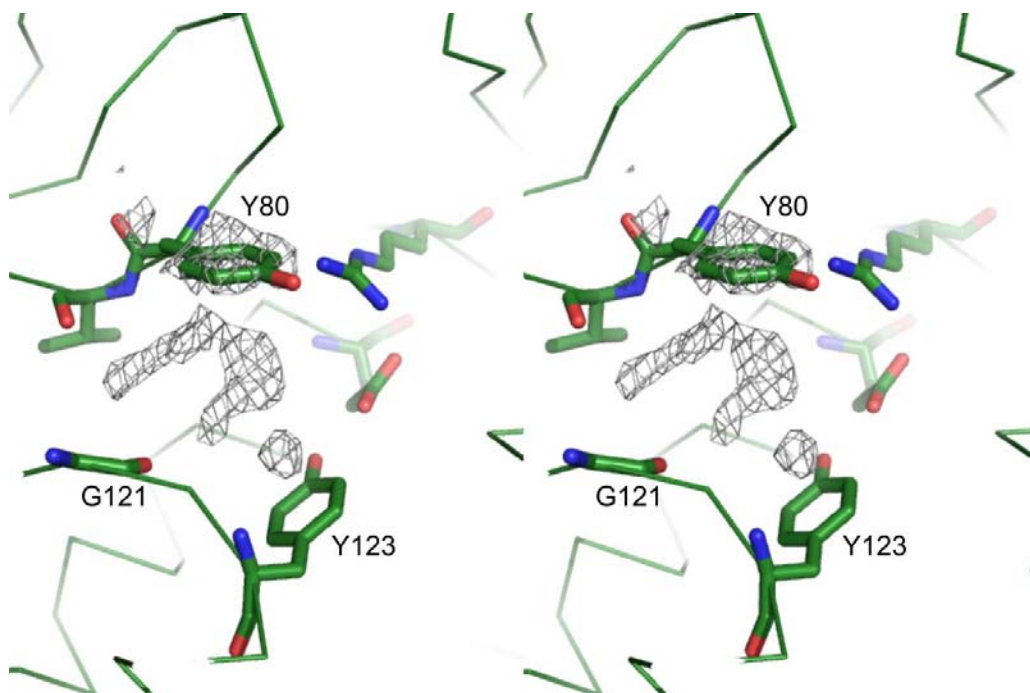


Figure 29. Stereo view of omit difference density for the RTA•5225540 complex. Tyr80 was mutated to Ala in the protein model, which was used to calculate the difference map. The map is contoured at the active site in grey mesh.

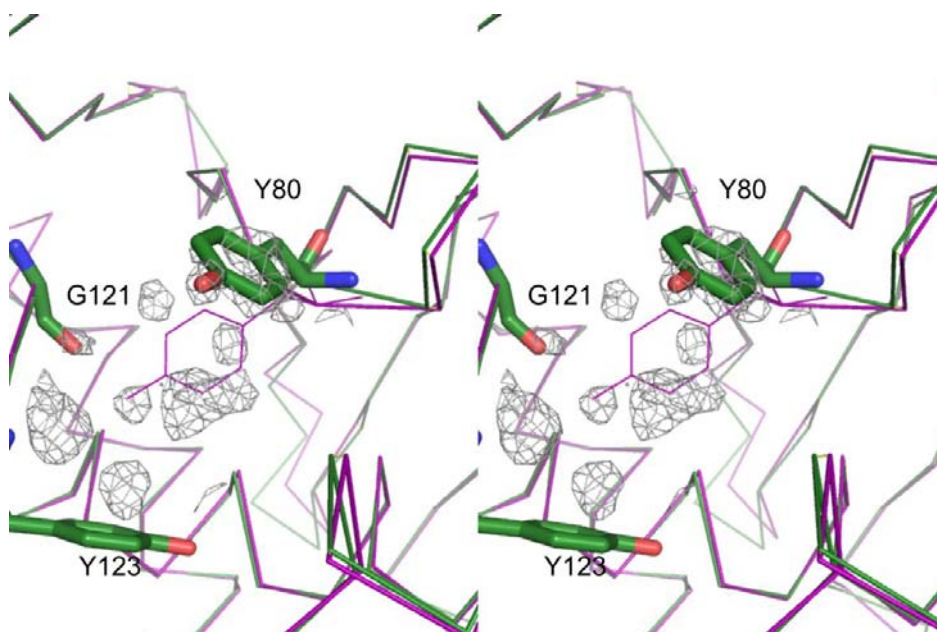


Figure 30. The stereo view of the superposition of RTA•7543758 complex and apo RTA (1RTC.pdb). The active side residues of RTA•7543758 complex are shown in stick (green), and the apo protein is shown in purple.

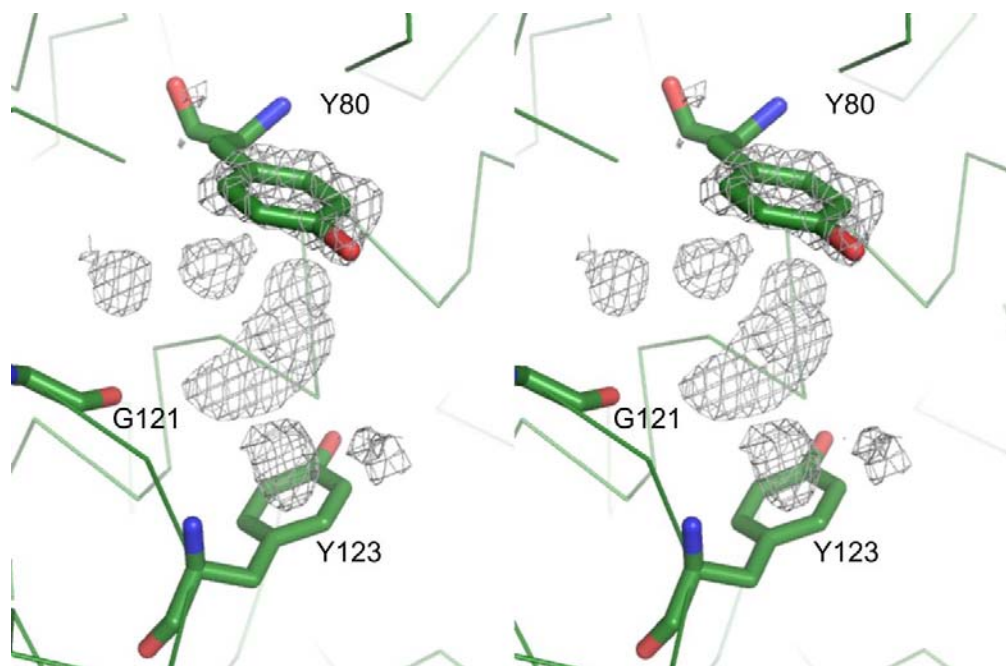


Figure 31. The stereo view of omitted difference density of the RTA•7780474 complex. The map is contoured at the active site.

In this practice, we selected 306 to assay from about 50,000 compounds in a virtual screening. Of these, five compounds were identified as RTA inhibitors. The hit rate was 1.6%, which was 16 times higher than average hit rate of HTS (0.1%) from other studies (Roberts M. Stroud 2008). All hits identified were novel inhibitors in terms of structure, which expanded the scaffold variety available for drug development, and will provide new information for inhibitor design once the exact binding modes for the complexes are revealed by crystallography. We hoped that some small ligands for the secondary pocket could be found from Chembridge Fragment Library, however all the hits from all the ChemBridge Libraries targeted only the specific pocket. The ChemBridge Fragment Library is a very small set for virtual screening, containing only 4000 compounds. We might need to screen larger libraries in order to find hits for the second pocket. In general, the VDS strategy was successful in identification new RTA and SItA inhibitors, which inspired us to continue on this track to ultimately find satisfactory drug leads.

Chapter 4: High throughput screening methods

4.1 INTRODUCTION

Although the virtual screening method has been explored for the RTA inhibitor search, as discussed previously, we also explored the utility of high throughput screening, HTS, as an alternative way to identify inhibitors. This was necessary since none of the docking programs proved terribly reliable in identifying potential hits from screening libraries, based on their scoring performance (I. Muegge 2001; Taylor, Jewsbury et al. 2002; Roberts M. Stroud 2008). The HTS had two main thrusts, one a cell based screen and the other an *in vitro* screen.

The cell based HTS was carried out by our collaborator Dr. Nicholas Mantis at the Wadsworth Center (NYS Department of Health). His group screened compound libraries at the NSRB (National Screening Laboratory for the Regional Centers of Excellence in Bio-defense and Emerging Infectious Diseases). They used a Vero cell-based HTS assay, by which they screened 29 libraries, containing over 200,000 compounds. . The cell-based assay is relatively fast and economical, but it is complex and may not reflect direct interactions between compounds and the RTA protein target. That is, the cellular intoxication pathway may be disrupted at any point in a complicated uptake pathway, which may produce false positives for our program. In addition, cell-based assays require special treatment to prevent contamination that can affect reproducibility; this would also produce false positives. For this reason, it is imperative to have a good cell-free HTS assay to establish RTA as the target and to quantify the interaction. The assay should ideally be simple, robust, reproducible, and economical. We adapted our *in vitro* translation assay to run on

a robotic system as a HTS assay. The resulting assay was workable but had several drawbacks. It was expensive and error prone. In order to have a robust assay for both a stand alone HTS and for testing compounds selected from other screens, we made considerable efforts on new assay development, which are discussed in this chapter, along with results of the HTS.

4.2 MATERIALS AND METHODS

4.2.1 Cell-based assay (Wadsworth Center)

100 μ l of Vero cells (50000 cells/ml) (DEME+10% FBS) were seeded into a 96-well plate and incubated at 37 °C for 15 hr in the presence of 5% CO₂. 50 μ l of medium (DEME+10% FBS) containing ~100 μ M of screening compound was added to the plate before incubation at 37 °C for one hour. Another 50 μ l of medium containing 0.3 nM ricin toxin was added to compound-containing wells and the plate was incubated at 37 °C for ~24 hr. For the positive control, 100 μ l of medium without compound or ricin toxin was added to the Vero cells, while the negative control contained ricin toxin only, without test compounds. Then, 50 μ l of growth medium, containing the dissolved compounds and ricin, were transferred to and mixed with 50 μ l of Gelltiter-Glo, which contained luciferase and luciferin. Luminescence was generated in the presence of ATP, and serves as an indicator of metabolically active cells. Finally, the percentage of cell viability of each well was calculated by comparing luminescence with that of the control.

4.2.2 Cell-free HTS for RTA inhibitor screening

The whole process of HTS includes three steps: *in vitro* translation producing luciferase, engaging the luciferase to produce light, and luminescence

reading. Each step may have several sub steps. The broad outline of the assay was shown previously as Figure 11 in Chapter 2.

Modifying our bench top translation assay to use in a multi-well format was surprisingly difficult, but with effort was made to work reasonably well. Setting up the plate assay is complicated; the current method is outlined in Figure 32.

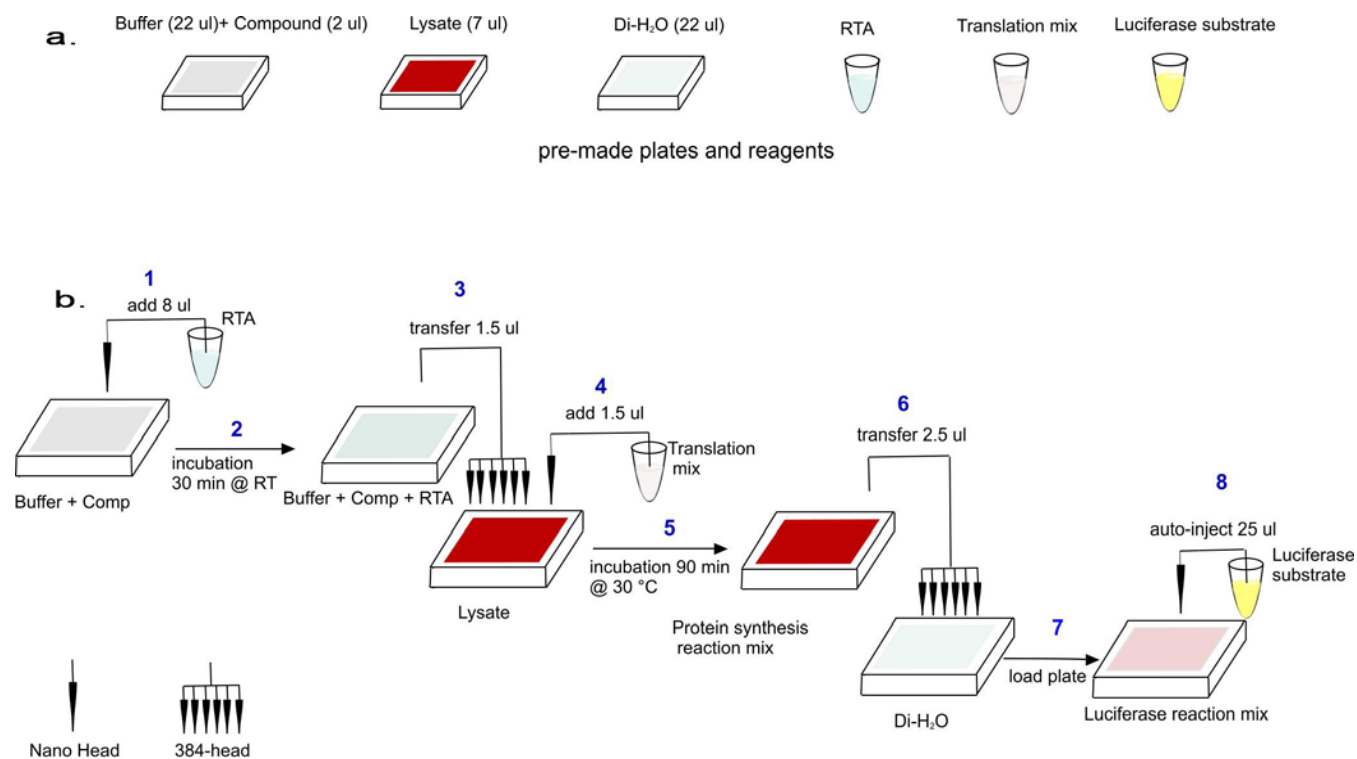


Figure 32 The *in vitro* HTS assay. a. preparation prior to the assay: buffer plate, containing 22 μ l HEPES buffer (20 mM, pH 7.6) and 2 μ l of screening compound stock; pre-aliquot lysate plate (7 μ l); Di-H₂O plate (22 μ l); RTA dilution (10 nM); translation mix made from components provided by Promega; and luciferase substrate reagent (Promega). Note, it was not necessary to make the Di-H₂O plate (22 μ l) and luciferase substrate reagent prior to the assay; they can be prepared during 90 min incubation instead.

b. HTS assay procedure: (1). Add 8 μ l prepared RTA dilution into each well of the pre-made buffer/compound plate using the Phoenix robot nano head. (2). Incubate 30 min at room temperature. (3) Transfer 1.5 μ l mixture of (2) into lysate plate (Inhibition controls were manually added into the lysate plate). (4) Immediately initiate the translation reaction by adding 1.5 μ l supermix using the nano head after (3). (5) Seal the plate and incubate at 30 °C for 90 min. (6) Transfer 2.5 μ l of the reaction mix to a water plate. (7) Load the plate into the plate reader (Envision 2013, Perkin Elmer). (8) Auto-dispense 25 μ l of luciferase substrate reagent into each well of the plate.

In each high-throughput screen we used three controls: a positive control containing dilute amounts of DMSO and KOH, the RTA inhibition control containing the known inhibitor pterioic acid (PTA supplied from Sigma) and RTA with DMSO/KOH, and finally a negative control containing RTA with DMSO/KOH. The final concentrations of components in these controls, in a 10 μ l total reaction volume were: 0.94% DMSO/1.4 mM KOH in HEPES for the positive control, 375 μ M PTA/6.25 nM RTA/0.94% DMSO/1.4 mM KOH for the RTA inhibition control and 6.25nM RTA/0.94% DMSO/1.4mM KOH for the negative control.

RTA protein solution after column purification is usually at 0.5~0.7mg/ml. Fresh protein was typically serially diluted 1000 fold (3 dilutions, 10 fold/dilution) into 20mM HEPES (pH 7.6). A dose response of the enzyme was performed prior to the HTS screen to assess activity and establish the exact dilution required to give about 95% translation inhibition. 8 μ l RTA was then added into a pre-made buffer/compound plate. That plate contained 22 μ l 20mM HEPES pH 7.6 and 2 μ l screening compound solution (~10 mM in 100% DMSO, or 1 μ l 100% DMSO for the negative control wells). The buffer/compound plate was prepared by Dr. Eun Jeong Cho in the TI3D core facility using the Janas (Perkin Elmer) automated liquid handling system. For the inhibition control, an equivalent amount of RTA/PTA was manually added into the appropriate wells. Then, the plate was incubated at room temp for 30 minutes prior to being added to the reaction. A short centrifugation was performed after each addition to assure mixing.

A volume of 7 μ l of nuclease treated rabbit reticulocyte lysate mix (Promega) was pre-aliquoted into a 384-well assay plate using an automated liquid handling system, Phoenix (Art Robins Instruments). The reaction

“supermix” was then assembled on ice using the Promega supplied components (150 μ l amino acids stock, 150 μ l RNasin RNase inhibitor stock, 300 μ l luciferase mRNA ~20 ng/ μ l stock, and 525 μ l nuclease-free water). The total volume of “supermix” was for 750 reactions so that the waste caused by priming nano head tubing was covered. The luciferase template used was mRNA provided by Promega in the kit as “Luciferase Control RNA.” Template RNA can also be synthesized using the RiboMAX Large Scale RNA Production System (Promega). Once the 30 minute RTA “co-incubation” with test compounds was finished, 1.5 μ l of this mixture was added to each well of the pre-dispensed lysate plate. 1.5 μ l of the translation supermix was finally added to the lysate/RTA/test compound to initiate the translation reaction. A short centrifugation at 600 rpm for 1 min was done for mixing the reaction solution. The reaction mixture was incubated at 30 °C for 1.5 hr, and stopped by incubating at -20 °C for at least 20min. Reaction mixtures can be stored at -20°C for up to two months or at -70°C for six months with no significant effect on luciferase activity.

To measure luciferase translation, 2.5 μ l of the completed translation assay mixture was transferred to the wells of an untreated, non-sterile, white 384-well plate (Nunc), which contained 22 μ l of Di-H₂O in order to increase accuracy of the transferring. After a short centrifugation, the plate was loaded onto a Perkin Elmer EnVision 2103 Multilabel Plate Reader and 25 μ l of Luciferase Substrate (Luciferase Substrate - Promega) was added to each well with the injector on the reader at a speed of 200 μ l/s. The plate was immediately read after the addition of the substrate. A single measurement time of 5 sec was used to measure the luminescence from each well. The Z' factor is a standard measure of the robustness of a HTS assay. It was calculated by the equation, $Z'=1-$

$(3*SD_p+3*SD_n)/(\mu_p-\mu_n)$. Here, SD stands for the standard deviation of the positive controls and negative controls respectively; μ presents the average value of the positive controls and negative controls respectively.

In order to test the effect of DMSO on the Rabbit Reticulocyte Lysate System, reactions were carried out at various DMSO concentrations in 15 μ l samples to determine the maximum percentage of DMSO that the assay could tolerate. DMSO concentrations tested were: 0, 0.033, 0.067, 0.1, 0.5, 1, 2, and 5% DMSO.

4.2.3 Circular Dichroism experiments

A Jasco 715 spectropolarimeter was used to acquire circular dichroism spectra and to record protein melting curve data. The instrument was equipped with a thermostated cell holder and a Neslab-111 circulating water bath. A quartz spectrophotometer cell with an optical path length of 1 mm was used (Starna Cells, Inc). Experiments were performed in 2.5 mM Phosphate (pH 6.5)/2.5 mM NaOH buffer condition (final concentration in cell).

4.2.4 Differential Scanning Fluorimetry (DSF) experiments

An Mx3005P QPCR system (Agilent Technologies) was used to monitor protein melting curves. 96-well polypropylene plates (natural, non sterile, Aligent) were used for sample reading. The total volume of each sample was prepared at 50 μ l, containing 4.8 μ M RTA, 500 μ M testing compounds (PTA, PBA, 7CP)/1% DMSO, and 5x dye (Sypro Orange) in HEPES (100 mM)/150 mM NaCl (pH 7.5) buffer. 20 μ l of the above mixture was transferred into plate wells, followed by a 1 min centrifugation at 600 rpm in a plate centrifuge (Hermle Z 300, Labnet). It is worth noting that the dye was mixed with the protein immediately before adding into the plate. Samples were kept on ice before loading. The temperature scan

was set up from 20 °C to 70 °C at a speed of one degree per min. The excitation wavelength was set at 492 nm and the emission wavelength was at 610 nm using internal filter sets.

4.2.5 Isothermal Titration Calorimetry (ITC) experiments

A MicroCal VP-ITC calorimeter was used to detect the direct binding between the ligand and RTA. The experiments were done in 20 mM HEPES pH 7.5/2.5 mM NaOH. RTA was dialyzed in the above buffer and adjusted to a final concentration of 50 µM. The ligand (7CP) was initially dissolved in 0.05 N NaOH and diluted by 20 mM HEPES (pH 7.5) to a final concentration of 1 mM in 20 mM HEPES pH 7.5/2.5 mM NaOH. All reagents were degassed under vacuum and filtered through a 0.2 µm polyethersulfone membrane using a Corning filter system. RTA was transferred into the calorimeter central chamber, and the ligand solution was added through the injector syringe. The injected volume was ~5-8 µl and the injection speed was 0.5 µl/sec.

4.3 RESULTS AND DISCUSSION

4.3.1 Hits from HTS assays

4.3.1.1 Hits from cell-based HTS assay

Our collaborator, Dr. Mantis, screened a total 29 compound libraries at the National Screening Laboratory for the Regional Centers of Excellence in Biodefense and Emerging Infectious Diseases (NSRB). The libraries contained over 200,000 compounds and were tested using a cell-based HTS assay. Figure 33 is a cartoon that simply describes how the assay was done. Table 9 shows the libraries that have been screened. This assay clearly discriminated between ricin treated versus untreated cells (or cells protected by a compound) as

evidenced by Z-prime factors in the range of 0.7-0.9. The Z-prime factor is an indicator of the robustness of an HTS assay, and values greater than 0.7 are considered good, whereas ≥ 0.9 are excellent (Zhang, Chung et al. 1999). All compounds in the library were tested in a primary screen, and then ~200 hits from the primary screen were retested in a secondary screen in order to confirm the activity. This was followed by a tertiary screen to test validated hits in a dose depended matter.

All the hits that were confirmed in cell-based assay were then tested in our laboratory using our *in vitro* translation assay to confirm that RTA was the inhibitor target. Twenty three compounds were identified that were able to preserve over 80% cell viability in a ricin challenge. We subsequently showed that eight compounds were inhibitory to ricin A chain, and 5 compounds inhibited Slt A chain in the *in vitro* translation assay (Table 10).

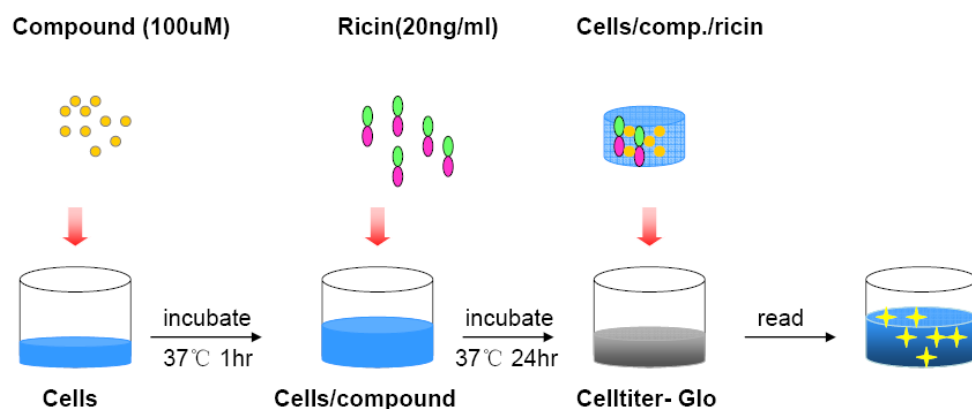


Figure 33. Cartoon of the cell-based assay for ricin inhibition.

Library name	Number of comp.	Library name	Number of comp.
Biomol ICCB	480	ChemDiv3	11616
Ninds CC 2	1040	Maybridge4	4576
Prestwick 1	1120	Biomol-TT 1	8518
ICBG 9 - fungal	1760	Bionet 2	1700
Asinex 1	12378	ChemDiv2	8560
ChemBridge 3	10560	Enamine 1	6004
ChemDiv4	14677	I.F. Lab 2	292
Enamine 2	26576	Maybridge 2	704
Life chem 1	3893	Maybridge 3	7639
Maybridge 5	3212	Peakdale 2	352
ChemDiv3	4928	Bionet 1	4800
Maybridge 4	352	CEREP	4800
Maybridge 2	352	ChemDiv 1	28864
Maybridge 1	8800	Peakdale 1	2816
ChemDiv 6	18656	Total	200025

Table 9. Libraries screened by cell-based HTS at the NSRB

	Structure	CID	in vivo IC50 (uM)	in vitro IC50 (uM)-Slt	In vitro IC50 (uM)-RTA
1*		2399297	303	40	30
2		2391135 dev.	UN	500	125
3		1300461	UN	250	150
4*		2977768	ND	> 500	150
5		2796234	ND	slight	300
6		1192842	ND	slight	400
7		2391135	ND	NO	500
8		7151481	ND	NO	500
9*		16271106	133	> 500	1000
10		2796234 dev.	ND	slight	1400
11		2084634	28	NO	slight inhibition w/BSA

Table 10. Hits selected from the cell-based HTS. Compounds 1, 4 and 9 are labeled with a “*”, indicating they exhibited slight cell toxicity. Compounds 2 and 10 are derivatives identified by our similarity search, not from the cell-based HTS.

The *in vitro* IC₅₀s of these RTA inhibitory compounds were within a range of high micromolar to low millimolar levels. Compound 1, which also exhibited slight cell toxicity, was the most potent one with an IC₅₀ of 30 μ M and 40 μ M against RTA and SlitA respectively (*in vitro*). Besides compound 1, compounds 4 and 9 also showed some cell toxicity (labeled with a “star” in table 10). Most hits showed different inhibitory effects on RTA versus SlitA, especially the relatively weaker RTA inhibitors, such as compound 5, 6, 7, and 8. In particular, compounds 7 and 8 didn’t inhibit SlitA at all in the *in vitro* assay. The selectivity of these hits suggested that some of the compounds might occupy, at least partially, the second pocket of RTA. The SlitA structure has a secondary pocket shallower than that of RTA, although the catalytic sites of both proteins are essentially identical (Fraser, Chernaia et al. 1994). Crystallization of these hits complexed with RTA is ongoing in our laboratory; the crystallization methods were discussed previously in this dissertation.

Most of the compounds identified as RTA inhibitors in this joint cell and *in vitro* system have very different structures from our previously known RTA inhibitors. Some of the new inhibitors fall into novel structural categories. For example, compounds 3, 5, 6 and 10, share a common sulfonamide containing pendant group that branches out from different aromatic ring structures. This may indicate there are important properties about this moiety with respect to their inhibitory effect. This may also provide a starting point to find better RTA inhibitors by similarity searches, a method we used successfully to find a better derivative for compound 7 in terms of binding affinity to RTA and SlitA.

Compound 7 was originally found in the cell screen and confirmed *in vitro* to have an IC₅₀ of 500 μ M against RTA, although it had no measurable inhibitory

effect on SltA. By similarity search, I identified a derivative, compound 2, with an IC_{50} of 125 μ M against RTA and 500 μ M against SltA. The substitution of one cyclohexyl group by a fluorophenyl group gave a stronger binding between the ligand and the protein. This suggested that the binding site for this group might need an aromatic stacking force provided by the aromatic phenyl, but not available with cyclohexane. The compound was sent to the Wadsworth Center and tested in the cell assay. The results (Figure 34) showed that the compound maintained about 50% cell viability at 28 μ M without any cell toxicity observed. However, above this concentration, the compound showed increasing cell toxicity with the increase of concentration. About 50% of the cells remained alive when the compound concentration reached 112 μ M, with or without ricin, suggesting that ricin activity might be inhibited completely by the compound at this concentration. However, it was not known, based on current experiments, why a substitution of the cyclohexyl group by a fluorophenyl group would introduce the cell toxicity. In order to solve this problem, more analogs of compound 7 should be tested to provide instructive information on structure and activity relationships.

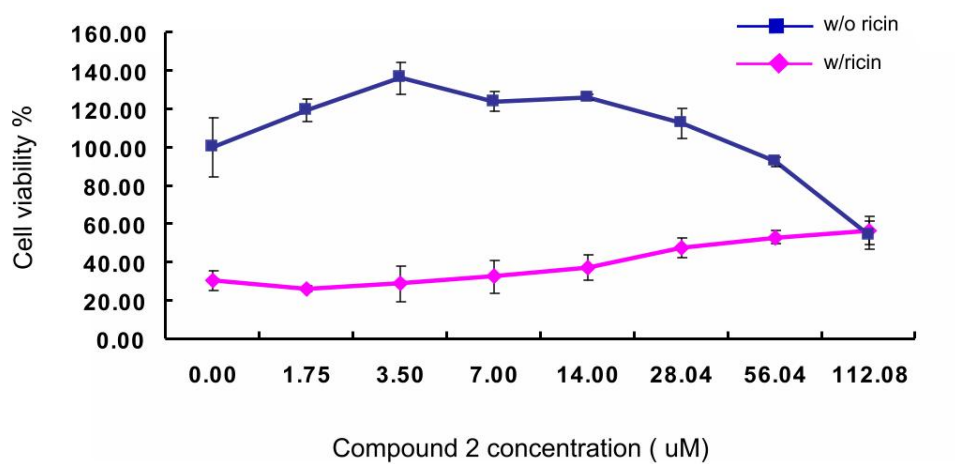


Figure 34. Compound 2 dose response curves in the cell-based assay. The purple line illustrates the dose response of compound 2 in the presence of ricin. The blue line depicts the effect of compound 2 by itself on the cell viability in a dose-dependant manner.

It is also worth noting that a number of hits reported from the cell-based HTS assay turned out to be false positives. We found one hit, for example, that was a strong positive regulator of protein translation and suppressed ricin intoxication in this way. Another intriguing case is that of compound 11 (Table 10), which showed relatively strong inhibitory effects in the cell-based assay with an IC_{50} of 28 μ M. It also appears to inhibit RTA in an assay devoid of BSA. However, it showed no activity in our *in vitro* assay in the presence of BSA. Compound 11 is an example of what are called non-specific inhibitors (McGovern et al., 2002; Seidler et al., 2003; Feng et al., 2005; Feng and Shoichet 2006). At high concentration, these compounds self-aggregate into a suspension of large particles that nonspecifically associate with, or adsorb proteins and thereby inactivate them. Several agents can be used to block this nondiscriminatory sequestering, such as non-ionic detergents (0.01-0.1%) or the addition of 1mg/ml BSA. The proposed blocking mechanism by BSA was that it can saturate these aggregates prior to encountering the enzyme rather than disrupting the aggregate formation (McGovern, Caselli et al. 2002). During the confirmation of these cell-based hits by *in vitro* translation assay, we used 1mg/ml BSA to block the nonspecific binding discussed above, although 0.06% triton showed the same function as BSA. The eight hits listed in Table 10 were confirmed in the presence of BSA.

In conclusion, the cell-based HTS identified several hundred “hits” in the initial screen. Subsequent re-screening reduced that number, and the *in vitro* test for RTA inhibition reduced the number still further. Once false positives were removed, it appears that the real hit rate was only 0.004%, which was, interestingly, much lower than the VDS discussed previously. Even so, some

novel chemical platforms were identified and may prove useful in future drug discovery.

4.3.1.2 NIH clinical collection screening

As described above, we implemented the *in vitro* protein synthesis assay to a robotic system, using the Phoenix robot (Art Robins Instruments) as a cell-free HTS assay tool. All the screening libraries acquired by TI-3D have the test compounds, generally at a nominal concentration of 10 mM, dissolved in 100 % DMSO. Our tests showed that RTA, and the translation system, retain activity at 1 % DMSO. The assays are set up so that the final DMSO concentration is 1%, and the test compounds at a nominal concentration of 100 μ M. The final reaction volume was 10 μ l, and protein synthesis was maintained at about 60%-70% compared to samples without DMSO.

As described in the method section, the assay was done in a 384-well plate format. Figure 35 shows the layout of a plate with 8 positive controls, just the protein translation system without either RTA or compound, 8 negative controls with RTA only, and 8 inhibition controls that contained RTA and 0.3 mM PTA, which usually inhibits 50% of RTA activity. The Z' factor obtained for the assay was typically 0.5-0.6, indicating the assay quality was just acceptable in terms of criteria for HTS. We screened a small library, the NIH clinical collection, containing 446 compounds using this assay system. This collection contains compounds that are known drugs or are in clinical trials. They are all non-toxic, drug like, and fairly soluble. Unfortunately, only one hit was initially identified after two duplicate HTS screens. However, this hit turned out to be a false positive that inactivated RTA by forming aggregates that adsorbed RTA. Figure 36 showed

the compound structure and the dose-dependent curve in the absence of BSA. In presence of BSA, the inhibitory activity was almost completely abolished.

We only screened the one library using this HTS method, because we realized that the assay had several major drawbacks. First of all, it was very expensive in the high throughput mode, costing over \$1/sample in reagents alone. In order to reduce the cost even to this level, we had scaled down the reaction volume to 10 μ l, but this resulted in another problem, aspiration and dispensing accuracy. This assay is a complex system, consisting of several enzymatic reactions, and therefore even a tiny error in volume could cause a large error in the final results. Secondly, in this assay, we used a lysate mixture, containing ribosomal proteins, initiation and elongation factors, and RNAs. With so many interacting components, we can not rule out the possibility that the test effects resulted from the interaction of the inhibitor candidate compounds with ribosomes or other factors (Too et al., 2009). Finally, the assay material, rabbit reticulocyte lysate, is very viscous and bubbly, and therefore it is a real challenge to use the automated liquid handling system to distribute the lysate with high accuracy. For all these reasons, we tried other methods in order to find a better assay for HTS, or at least for secondary screening, which are discussed in the following sections.

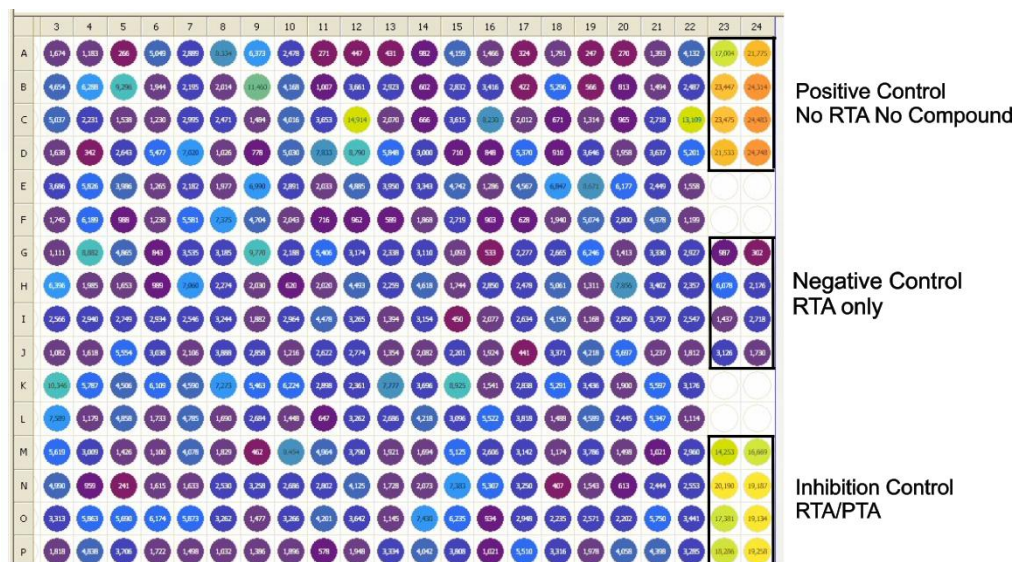


Figure 35. 384-plate layout for NCC library HTS.

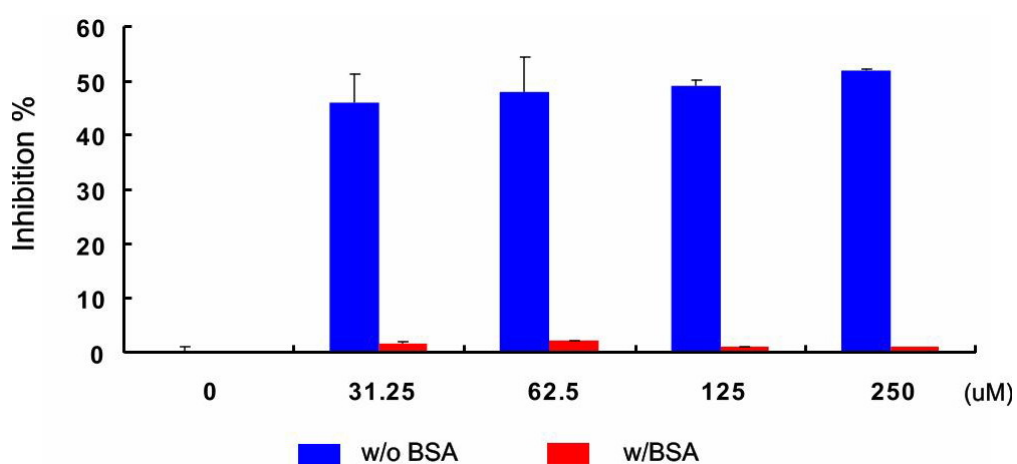
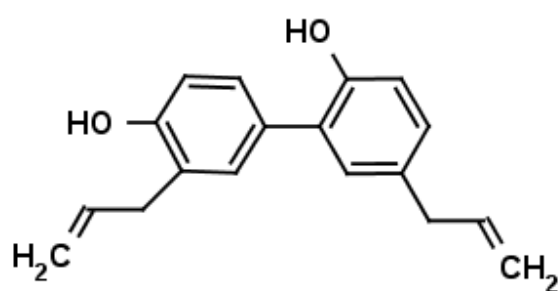


Figure 36. The honokiol structure and activity with and without BSA. The x-axis is the concentration of honokiol. Observed inhibition effects of honokiol against RTA are shown in blue and red with or without 1mg/m BSA respectively.

4.3.2 Binding detection by CD experiments

CD was used by others to detect adenine binding in RTA by monitoring the change of CD spectra in the far UV (near 220 nm) (T.S. Ramalingam 1993). However, the change was so subtle that it could have been caused by non-binding related factors, such as buffers or even just the difference between runs. However, CD still can be used to show binding by monitoring the change of protein unfolding process in the presence of the ligand. A protein can be stabilized upon binding a ligand, and therefore the melting temperature (T_m) of the protein may increase. In order to evaluate if there is a true correlation between binding and stabilization of protein, we ran a number of CD experiments to monitor RTA T_m change in presence of known ligands.

The CD full spectra were obtained first, which showed a distinct helix spectrum with a strong peak around 207 nm. Since many currently known RTA ligands require 0.05 N NaOH as solvent, we used 2.5 mM NaOH in the experiment buffer system. The CD spectra were essentially identical with or without NaOH. Figure 37a shows the CD spectra from 30 nm to 190 nm, for RTA with buffer and different compounds. Only PTA changed the RTA spectrum shape due to the absorbance of PTA in the far UV region. However, the change was negligible at 207 nm, which was the wavelength setup for T_m measurement.

Figure 37b shows the T_m shifts caused by buffer conditions and by addition of different compounds. We observed a decrease of the RTA T_m by about 3-4 degrees upon the addition of NaOH, compared to RTA in phosphate buffer. Since all T_m measured for RTA and RTA complexes were tested in presence of the same amount NaOH, the change of RTA T_m caused by NaOH was systematic and can be ignored. The experiments showed that RTA can be

stabilized by most of the known ligands tested. PTA showed the largest stabilizing effect on RTA, with a T_m increase of 4.45 °C, followed by 7CP with a increase of 2.93° at a concentration of 125 µM. PBA showed a slight stabilizing effect at 125 µM, but the effect was magnified with the increase of concentration by two fold.

R121 was one of PBA derivatives discussed previously. It did not show measureable RTA inhibition at 125 µM, however, it stabilized RTA by 3.73 °C when the concentration was increased to 625 µM. It appears that CD experiments are able to identify very weak binders that can't be detected in our normal activity assay. CD monitoring of the protein structure is apparently more sensitive than measuring the protein synthesis changes resulting from ligand binding.

As mentioned previously, it is possible that some compounds can form aggregates at high concentration and nonspecifically adsorb RTA, and might therefore they stabilize the protein. In order to test this hypothesis, we tested RTA T_m change in the presence of 125 µM Epigallocatechin gallate (EGCG), a well-known nonspecific binding compound. The results showed that the T_m of RTA decreased about 5 °C. This observation suggested that ligands eliciting an increase in T_m are likely to be true ligands. However, not all ligands necessarily increase the T_m . A known ligand, PPO, also destabilized RTA by 1.4 °C. The T_m data need to be viewed carefully in drug design..

Based on these observations, a CD experiment may be used as a method to provide supporting evidence for ligand binding, but it can not be used in a high throughput mode. In addition, the instrument can barely tolerate DMSO, even at 0.1%, which may limit its utility for testing compounds from commercial libraries.

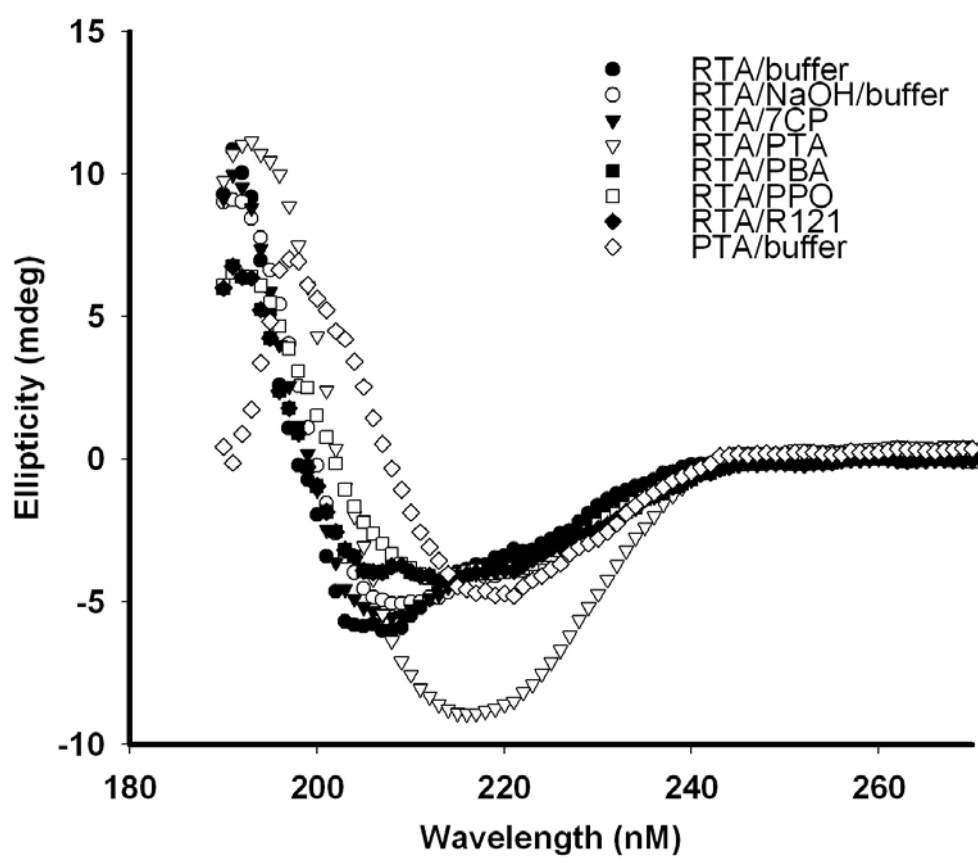


Figure 37a. The CD wavelength scan for RTA, and RTA with different compounds.

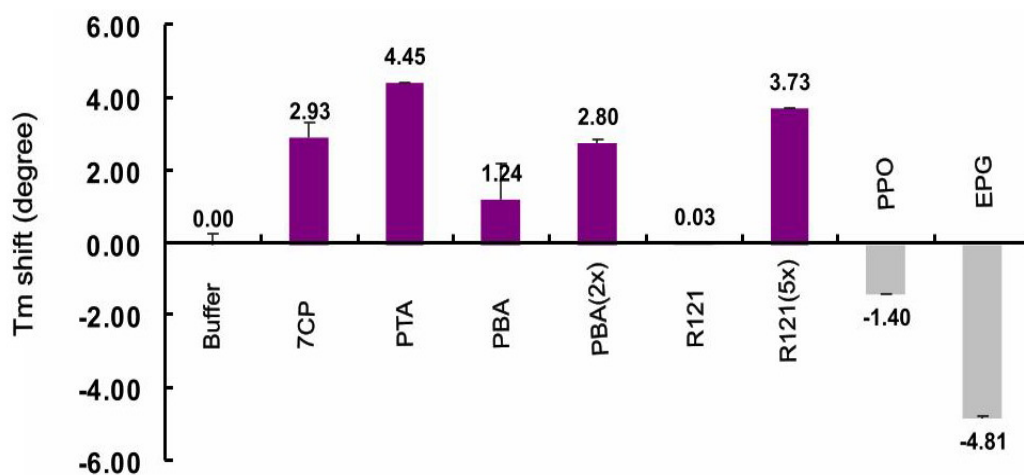


Figure 37b. RTA T_m shifts caused by test compounds. The compounds that stabilized RTA are shown in purple, and those that destabilized RTA are shown in grey. All compounds were tested at 125 μM with the exception of PBA and R121. PBA was tested at 125 μM and 250 μM (2x), R121 was tested at 125 μM and 625 μM (5x). In general, compounds known to bind RTA from kinetic data and X-ray structures raise the T_m of RTA several degrees.

4.3.3 Results from Differential Scanning Fluorimetry (DSF)

Differential Scanning Fluorimetry (DSF) also measures the T_m of proteins and detects ligand binding by a change in T_m (Niesen et al., 2007). In this assay, we measure the change in protein T_m in the presence of test compounds, but unlike the CD experiments, this method uses a fluorescent dye as reporter. The fluorescence intensity will increase when the dye nonspecifically binds to unfolded protein by hydrophobic interaction. One advantage of this method is that it can be readily implemented in a high throughput fashion. Figure 38 shows the results obtained from our preliminary tests. The assay can tolerate 2% DMSO, a concentration at which RTA still remains well-folded (based on kinetic tests). RTA folding as a function of temperature was monitored in the presence of 7CP, PTA and PBA. Among these only 7CP showed more than a 2 °C increase of RTA T_m . Surprisingly, there was no measurable change in RTA T_m caused by PTA. In addition, the RTA melting curve in the presence of PTA showed very low fluorescence intensity compared with the control and other curves. However, the PTA curve by itself looked normal, shown in pink open triangle in figure 36. It seems that compounds with certain chemical properties, like PTA, might quench the fluorescence of the dye by interacting in an unspecified way.

In order to evaluate the DSF method further, more experiments need to be done, including assessing buffer conditions, protein concentration, and temperature change rate. Based on the results we obtained so far, we can not conclude if it will be a good assay for HTS.

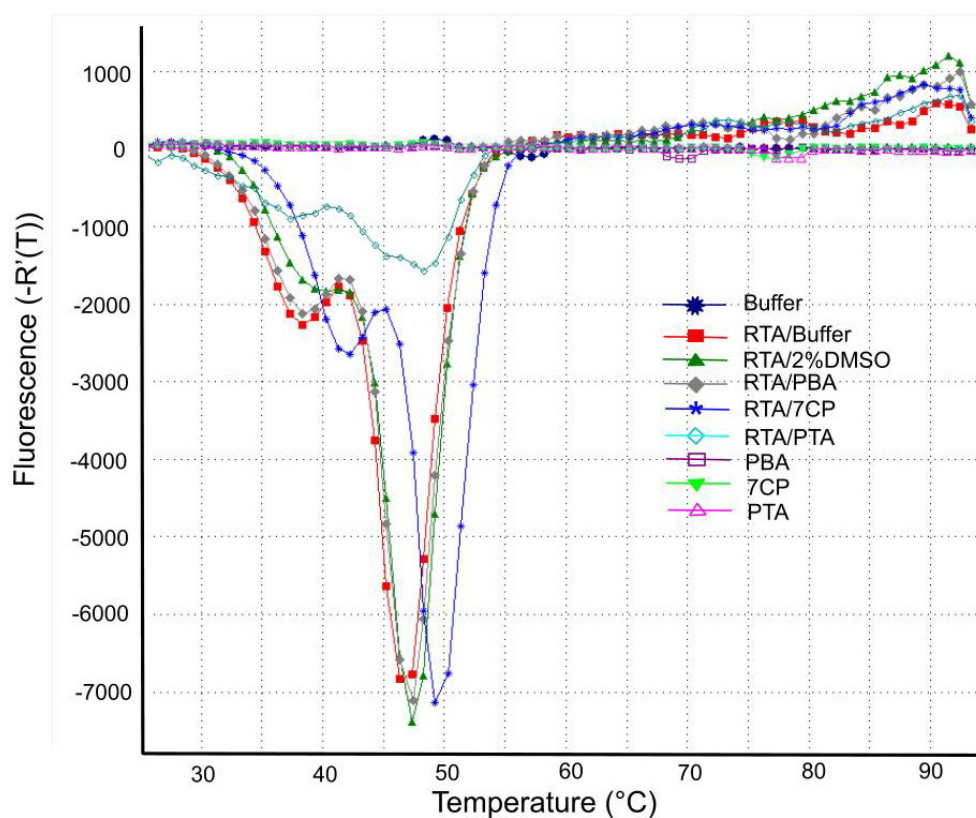


Figure 38. RTA T_m change by compounds. The x-axis is temperature in ($^{\circ}\text{C}$), and the y-axis is the first derivative of fluorescence of the dye. The base line curve and the curves for corresponding compounds were shown in different colors.

4.3.4 Results from Isothermal Titration Calorimetry (ITC)

ITC is a widely used biophysical method to detect ligand binding to proteins (Cooper 1999; Leavitt and Freire 2001; Turnbull and Daranas 2003; Weber and Salemme 2003). In this method the heat change is monitored while protein and ligand interact over time. Under many conditions, the method can measure true K_d values, and binding stoichiometries of the ligand. Unfortunately, this method requires a large amount of material, especially for weakly binding interactions. The specific concentration needed for protein and ligand depends on the titration order and K_d . For the binding partner that is injected into the cell chamber, the concentration needed is ideally 10 to 100 times of K_d in order to achieve a C values of 10~100. The C value is a parameter defined by Wiseman (Wiseman, Williston et al. 1989) to help clarify the conditions required to fit the isotherm binding equation with high accuracy (Wiseman, Williston et al. 1989; Indyk and Fisher 1998). The shape of the isotherm binding curve is dependent on the value of C. For C-values greater than 10, the curve is generally sigmoidal and the key parameters K_a and stoichiometry are easily defined. However, as C decreases below 10, the inflection point becomes poorly defined and the binding stoichiometry can no longer be determined experimentally. High C value may be easily obtained for tight binding reactions, but it is essentially impossible to meet this criterion for ligands binding with the value of K_d between 0.1 M and 1 M, which is unfortunately the range of binding affinities of most current RTA inhibitors.

Nonetheless, ITC is still very useful in providing relatively accurate K_d values for weak bindings under certain circumstances, which are (1) an adequate portion of the binding isotherm with good levels of signal-to-noise, (2) known

stoichiometry, and (3) receptor and ligand concentration known with high accuracy. The error in observed enthalpy change is highly dependent on the receptor concentration (the binding partner in cell chamber) for low C values, so the heat change might be off somewhat, but K_d provided by the analysis is still reasonably accurate (Turnbull and Daranas 2003).

The binding of RTA and its inhibitors belong in the low affinity classification, according to the IC_{50} values derived from the *in vitro* translation assay. RTA precipitates at concentrations over 100 μ M, so the C value that can be achieved would probably be one or less for such ligands. However, the requirements discussed above can still be met in some cases and allow us to acquire accurate K_d values. For example, 7CP was used for the first ITC test during the validation process because of its relatively high solubility compared with other known RTA inhibitors. In the assay, about 1 mM 7CP solution was used to titrate RTA (50 μ M). The results showed that the binding K_d for 7CP was about 50 to 80 μ M when the fitting was done assuming a fixed stoichiometry of 1. Figure 38 illustrates one such data set.

The curve looks very smooth with no abnormal data points. The top panel of the figure is the raw data for the heat change versus time during the titration; it shows that RTA was not completely saturated by the end of titration. This is a common situation for low affinity systems with a C value below 1. Turnbull provided evidence that the binding K_d can still be calculated with high accuracy for low affinity system with saturation level above 70% as long as the criteria were met as discussed above (Turnbull and Daranas 2003). In particular, we knew that the stoichiometry was 1 from the X-ray data, and that concentrations of protein and ligand were accurately measured. As a result, the K_d calculated from

this data set was 80 μM , consistent with a saturation level of about 90% (which was also consistent with the raw data profile).

The lower panel of Figure 39 is the integrated enthalpy change for each injection versus the molar ratio of the ligand to the receptor. If the C value is within the ideal range, 10~100, the molar ratio at 50% saturation level will be the stoichiometry of binding. However, in this particular experiment, an accurate stoichiometry can not be obtained by curve fitting since the C value was beyond the ideal range, but we have assumed a value of 1, as described above, and this seems justified based on our X-ray structure.

The K_d obtained from ITC, 80 μM , was much lower than the IC_{50} , 300 μM , obtained from the activity assay. This observation can be explained by considering the equation of $\text{IC}_{50} = K_i (1 + [\text{S}]/K_m)$ (Cheng and Prusoff 1973). IC_{50} always represents an upper bound for K_d , and only when $[\text{S}] \ll K_m$, does IC_{50} approximate K_i . However, in our assay system, the substrate concentration in the translation system was unknown, so it is possible that the concentration of the ribosome substrate was higher than K_m . K_m for RTA has been accurately measured using *Artemia salina* ribosomes in a well controlled initial rate study as 0.4 μM (Ready et al., 1991) and estimated for rabbit reticulocytes at 0.1 μM (Olsnes et al., 1975). It is also worth noting that IC_{50} values are measured in an arbitrary way, allowing the active enzyme and translation system to run in parallel for a convenient time period; the values observed are highly dependent on the choice of assay parameters. An IC_{50} value can be halved by simply doubling the incubation time. The K_d value measured for 7CP by ITC is very encouraging, and suggests our inhibitors are stronger ligands than we had initially suspected.

Based on the above analysis, ITC could be used to detect direct binding even for a low affinity system, like RTA and its current ligands. ITC can provide the true binding affinity of ligands, which would be very difficult to assess with our kinetic assay system. As our program develops better and more specific inhibitors ITC should become even more important in our analyses.

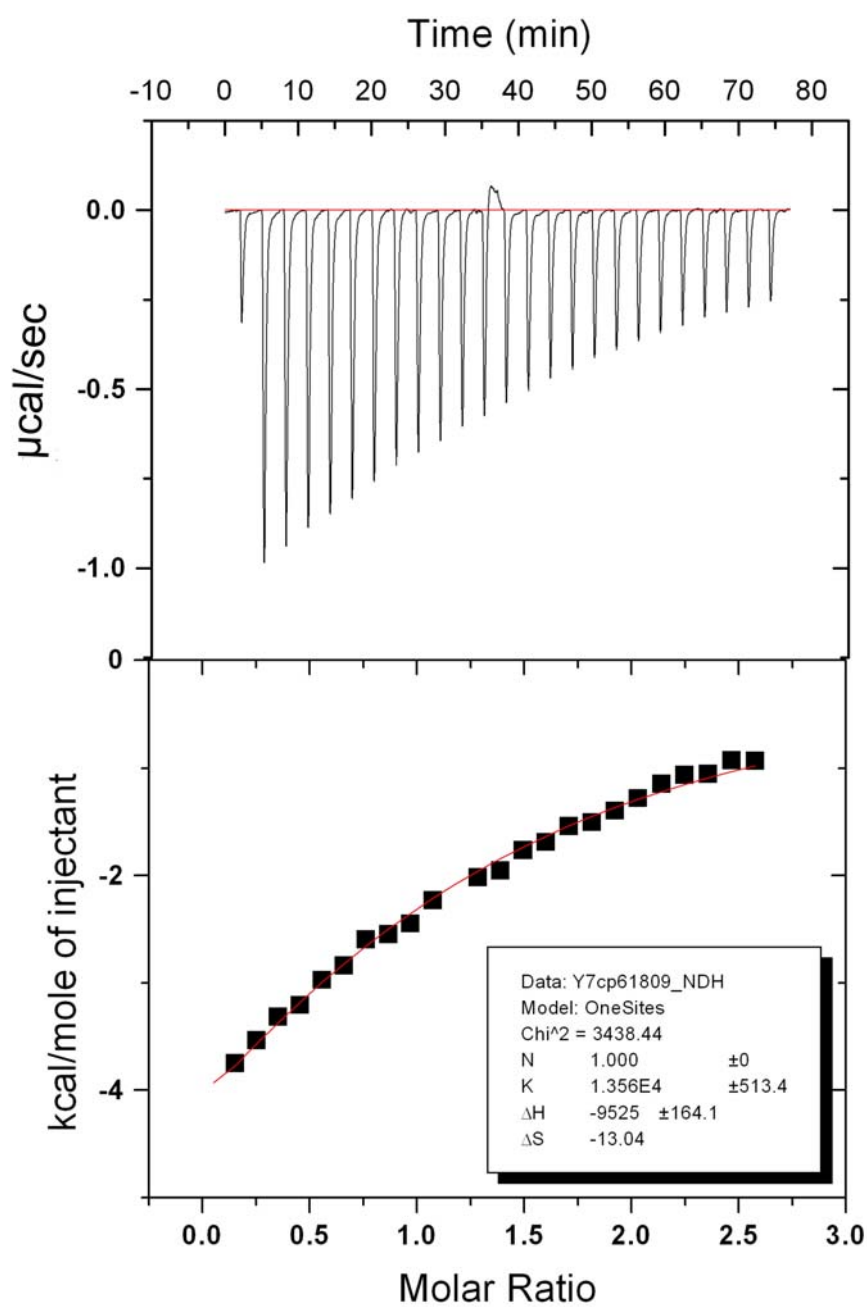


Figure 39. Representation of ITC data for the interaction of 7CP and RTA. The top depicts the heat change versus time. The bottom is the integrated data, with the best fitting lines in red.

References

- Abagyan, R. and M. Totrov (2001). "High-throughput docking for lead generation." Curr Opin Chem Biol **5**(4): 375-82.
- Appelt, K., R. J. Bacquet, et al. (1991). "Design of enzyme inhibitors using iterative protein crystallographic analysis." J Med Chem **34**(7): 1925-34.
- Argent, R. H., A. M. Parrott, et al. (2000). "Ribosome-mediated folding of partially unfolded ricin A-chain." J Biol Chem **275**(13): 9263-9.
- Bai, Y., A. F. Monzingo, et al. (2009). "The X-ray structure of ricin A chain with a novel inhibitor." Arch Biochem Biophys **483**(1): 23-8.
- Barbieri, L., M. G. Battelli, et al. (1993). "Ribosome-inactivating proteins from plants." Biochim Biophys Acta **1154**(3-4): 237-82.
- Barbieri, L., L. Polito, et al. (2006). "Ribosome-inactivating proteins in edible plants and purification and characterization of a new ribosome-inactivating protein from Cucurbita moschata." Biochim Biophys Acta **1760**(5): 783-92.
- Biot, C., H. Bauer, et al. (2004). "5-substituted tetrazoles as bioisosteres of carboxylic acids. Bioisosterism and mechanistic studies on glutathione reductase inhibitors as antimalarials." J Med Chem **47**(24): 5972-83.
- Butterworth, A. G. and J. M. Lord (1983). "Ricin and Ricinus communis agglutinin subunits are all derived from a single-size polypeptide precursor." Eur J Biochem **137**(1-2): 57-65.
- Calderwood, S. B., F. Auclair, et al. (1987). "Nucleotide sequence of the Shiga-like toxin genes of Escherichia coli." Proc Natl Acad Sci U S A **84**(13): 4364-8.
- Charifson, P. S., J. J. Corkery, et al. (1999). "Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins." J Med Chem **42**(25): 5100-9.
- Chaudhary, V. K., Y. Jinno, et al. (1990). "Pseudomonas exotoxin contains a specific sequence at the carboxyl terminus that is required for cytotoxicity." Proc Natl Acad Sci U S A **87**(1): 308-12.
- Chen, X. Y., T. M. Link, et al. (1998). "Ricin A-chain: kinetics, mechanism, and RNA stem-loop inhibitors." Biochemistry **37**(33): 11605-13.

- Cheng, Y. and W. H. Prusoff (1973). "Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction." Biochem Pharmacol **22**(23): 3099-108.
- Chung, C. W. (2007). "The use of biophysical methods increases success in obtaining liganded crystal structures." Acta Crystallogr D Biol Crystallogr **63**(Pt 1): 62-71.
- Cieplak, W., Jr., R. J. Messer, et al. (1995). "Role of a potential endoplasmic reticulum retention sequence (REDEL) and the Golgi complex in the cytotoxic activity of Escherichia coli heat-labile enterotoxin." Mol Microbiol **16**(4): 789-800.
- Coan, K. E. and B. K. Shoichet (2008). "Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors." J Am Chem Soc **130**(29): 9606-12.
- Cooper, A. (1999). "Thermodynamic analysis of biomolecular interactions." Curr Opin Chem Biol **3**(5): 557-63.
- Correll, C. C., A. Munishkin, et al. (1998). "Crystal structure of the ribosomal RNA domain essential for binding elongation factors." Proc Natl Acad Sci U S A **95**(23): 13436-41.
- Correll, C. C., I. G. Wool, et al. (1999). "The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution." J Mol Biol **292**(2): 275-87.
- Donohue-Rolfe, A., D. W. Acheson, et al. (1991). "Shiga toxin: purification, structure, and function." Rev Infect Dis **13 Suppl 4**: S293-7.
- Donohue-Rolfe, A., G. T. Keusch, et al. (1984). "Pathogenesis of Shigella diarrhea. IX. Simplified high yield purification of Shigella toxin and characterization of subunit composition and function by the use of subunit-specific monoclonal and polyclonal antibodies." J Exp Med **160**(6): 1767-81.
- Edward R. Zartler, M. J. S. (2008). "Fragment-Based Drug Discovery."
- Ellenhorn, M. J., Barceloux, D. G. (1988). "Diagnosis and Treatment." Medical Toxicology: 1225-1227.
- Ellgaard, L., M. Molinari, et al. (1999). "Setting the standards: quality control in the secretory pathway." Science **286**(5446): 1882-8.

- Endo, Y. and K. Tsurugi (1987). "RNA N-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes." J Biol Chem **262**(17): 8128-30.
- Ewing, T. J., S. Makino, et al. (2001). "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases." J Comput Aided Mol Des **15**(5): 411-28.
- Feng, B. Y., A. Shelat, et al. (2005). "High-throughput assays for promiscuous inhibitors." Nat Chem Biol **1**(3): 146-8.
- Ferrini, J. B., M. Martin, et al. (1995). "Expression of functional ricin B chain using the baculovirus system." Eur J Biochem **233**(3): 772-7.
- Fraser, M. E., M. M. Chernaia, et al. (1994). "Crystal structure of the holotoxin from *Shigella dysenteriae* at 2.5 Å resolution." Nat Struct Biol **1**(1): 59-64.
- Fraser, M. E., M. Fujinaga, et al. (2004). "Structure of shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7." J Biol Chem **279**(26): 27511-7.
- Frigerio, L., N. A. Jolliffe, et al. (2001). "The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting." Plant Physiol **126**(1): 167-75.
- Girbes, T., J. M. Ferreras, et al. (2004). "Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria." Mini Rev Med Chem **4**(5): 461-76.
- Girod, A., B. Storrie, et al. (1999). "Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum." Nat Cell Biol **1**(7): 423-30.
- Glen Spraggon, P. V. (1996). "Progress in the Design of Immunomodulators Based on the Structure of Interleukin-1." Structure-Based Drug Design: 395-434.
- Gonatas, N. K., J. O. Gonatas, et al. (1998). "The involvement of the Golgi apparatus in the pathogenesis of amyotrophic lateral sclerosis, Alzheimer's disease, and ricin intoxication." Histochem Cell Biol **109**(5-6): 591-600.
- HARLEY, S. M. L., J. M. (1985). "In vitro endoproteolytic cleavage of castor bean lectin precursors." plant science **41**: 111-116.

- Hazes, B. and R. J. Read (1997). "Accumulating evidence suggests that several AB-toxins subvert the endoplasmic reticulum-associated protein degradation pathway to enter target cells." Biochemistry **36**(37): 11051-4.
- Hesselberth, J. R., D. Miller, et al. (2000). "In vitro selection of RNA molecules that inhibit the activity of ricin A-chain." J Biol Chem **275**(7): 4937-42.
- Hinshaw, J. E. and S. L. Schmid (1995). "Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding." Nature **374**(6518): 190-2.
- Holmberg, L. and O. Nygard (1996). "Depurination of A4256 in 28 S rRNA by the ribosome-inactivating proteins from barley and ricin results in different ribosome conformations." J Mol Biol **259**(1): 81-94.
- Holmes, D. S., R. C. Bethell, et al. (1993). "Synthesis and structure-activity relationships of a series of penicillin-derived HIV proteinase inhibitors containing a stereochemically unique peptide isostere." J Med Chem **36**(21): 3129-36.
- I. Muegge, M. R. (2001). "Small molecule docking and scoring." Reviews in Computational Chemistry **17**.
- Indyk, L. and H. F. Fisher (1998). "Theoretical aspects of isothermal titration calorimetry." Methods Enzymol **295**: 350-64.
- Ishizaki, T., C. Megumi, et al. (2002). "Accumulation of a 31-kDa glycoprotein in association with the expression of embryogenic potential by spinach callus in culture." Physiol Plant **114**(1): 109-115.
- Iversen, T. G., G. Skretting, et al. (2001). "Endosome to Golgi transport of ricin is independent of clathrin and of the Rab9- and Rab11-GTPases." Mol Biol Cell **12**(7): 2099-107.
- Jain, A. N. (2003). "Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine." J Med Chem **46**(4): 499-511.
- Jain, A. N. (2007). "Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search." J Comput Aided Mol Des **21**(5): 281-306.
- Jain, A. N. (2009). "Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation." J Comput Aided Mol Des **23**(6): 355-74.

- Jencks, W. P. (1981). "On the attribution and additivity of binding energies." Proc Natl Acad Sci U S A **78**(7): 4046-4050.
- Jones, G., P. Willett, et al. (1995). "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation." J Mol Biol **245**(1): 43-53.
- Jones, G., P. Willett, et al. (1997). "Development and validation of a genetic algorithm for flexible docking." J Mol Biol **267**(3): 727-48.
- Juan Alvarez, B. S. (2005). "Virtual Screening in Drug Discovery."
- Karmali, M. A., M. Petric, et al. (1985). "The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*." J Infect Dis **151**(5): 775-82.
- Keusch, G. T. and M. Jacewicz (1977). "Pathogenesis of *Shigella* diarrhea. VII. Evidence for a cell membrane toxin receptor involving beta1 leads to 4-linked N-acetyl-D-glucosamine oligomers." J Exp Med **146**(2): 535-46.
- Kim, Y., D. Mlsna, et al. (1992). "Structure of a ricin mutant showing rescue of activity by a noncatalytic residue." Biochemistry **31**(12): 3294-6.
- Kim, Y. and J. D. Robertus (1992). "Analysis of several key active site residues of ricin A chain by mutagenesis and X-ray crystallography." Protein Eng **5**(8): 775-9.
- Koehler, R. T. and H. O. Villar (2000). "Statistical relationships among docking scores for different protein binding sites." J Comput Aided Mol Des **14**(1): 23-37.
- Lam, P. Y., P. K. Jadhav, et al. (1994). "Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors." Science **263**(5145): 380-4.
- Lamb, F. I., L. M. Roberts, et al. (1985). "Nucleotide sequence of cloned cDNA coding for preproricin." Eur J Biochem **148**(2): 265-70.
- Leavitt, S. and E. Freire (2001). "Direct measurement of protein binding energetics by isothermal titration calorimetry." Curr Opin Struct Biol **11**(5): 560-6.
- Liebman, S. W., Y. O. Chernoff, et al. (1995). "The accuracy center of a eukaryotic ribosome." Biochem Cell Biol **73**(11-12): 1141-9.

- Lipinski, C. A., F. Lombardo, et al. (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Adv Drug Deliv Rev **46**(1-3): 3-26.
- Llorente, A., A. Rapak, et al. (1998). "Expression of mutant dynamin inhibits toxicity and transport of endocytosed ricin to the Golgi apparatus." J Cell Biol **140**(3): 553-63.
- Lombardi, D., T. Soldati, et al. (1993). "Rab9 functions in transport between late endosomes and the trans Golgi network." Embo J **12**(2): 677-82.
- Lord, J. M. (1985). "Precursors of ricin and Ricinus communis agglutinin. Glycosylation and processing during synthesis and intracellular transport." Eur J Biochem **146**(2): 411-6.
- Lord, J. M. (1985). "Synthesis and intracellular transport of lectin and storage protein precursors in endosperm from castor bean." Eur J Biochem **146**(2): 403-9.
- Lord, J. M., harley S.M. (1985). "Ricinus communis agglutinin B chain contains a fucosylated oligosaccharide side chain not present on ricin B chain." FEBS J letter **189**(1): 72-76.
- Lord, J. M., L. M. Roberts, et al. (1994). "Ricin: structure, mode of action, and some current applications." Faseb J **8**(2): 201-8.
- Majoul, I., D. Ferrari, et al. (1997). "Reduction of protein disulfide bonds in an oxidizing environment. The disulfide bridge of cholera toxin A-subunit is reduced in the endoplasmic reticulum." FEBS Lett **401**(2-3): 104-8.
- Marsden, C. J., D. C. Smith, et al. (2005). "Ricin: current understanding and prospects for an antiricin vaccine." Expert Rev Vaccines **4**(2): 229-37.
- Mason, J. S. and D. L. Cheney (2000). "Library design and virtual screening using multiple 4-point pharmacophore fingerprints." Pac Symp Biocomput: 576-87.
- McGovern, S. L., E. Caselli, et al. (2002). "A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening." J Med Chem **45**(8): 1712-22.
- McMartin, C. and R. S. Bohacek (1997). "QXP: powerful, rapid computer algorithms for structure-based drug design." J Comput Aided Mol Des **11**(4): 333-44.

- Michael A. Jarpe, C. H. P., Brian E. Szente, Howard M. Johnson (1996). "Structure and Functional Studies of Interferon: A Solid Foundation for Rational Design." Structure-Based Drug Design: 435-458.
- Miesenbock, G. and J. E. Rothman (1995). "The capacity to retrieve escaped ER proteins extends to the trans-most cisterna of the Golgi stack." J Cell Biol **129**(2): 309-19.
- Miller, D. J., K. Ravikumar, et al. (2002). "Structure-based design and characterization of novel platforms for ricin and shiga toxin inhibition." J Med Chem **45**(1): 90-8.
- Miller, M., J. Schneider, et al. (1989). "Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution." Science **246**(4934): 1149-52.
- Mlsna, D., A. F. Monzingo, et al. (1993). "Structure of recombinant ricin A chain at 2.3 Å." Protein Sci **2**(3): 429-35.
- Moazed, D., J. M. Robertson, et al. (1988). "Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA." Nature **334**(6180): 362-4.
- Montfort, W., J. E. Villafranca, et al. (1987). "The three-dimensional structure of ricin at 2.8 Å." J Biol Chem **262**(11): 5398-403.
- Monzingo, A. F., E. J. Collins, et al. (1993). "The 2.5 Å structure of pokeweed antiviral protein." J Mol Biol **233**(4): 705-15.
- Monzingo, A. F. and J. D. Robertus (1992). "X-ray analysis of substrate analogs in the ricin A-chain active site." J Mol Biol **227**(4): 1136-45.
- Murray, C. W. and M. L. Verdonk (2002). "The consequences of translational and rotational entropy lost by small molecules on binding to proteins." J Comput Aided Mol Des **16**(10): 741-53.
- Nilsson, T. and G. Warren (1994). "Retention and retrieval in the endoplasmic reticulum and the Golgi apparatus." Curr Opin Cell Biol **6**(4): 517-21.
- Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nat Protoc **2**(9): 2212-2221.
- Olsnes, S. and J. V. Kozlov (2001). "Ricin." Toxicon **39**(11): 1723-8.

- Olsnes, S., R. Reisbig, et al. (1981). "Subunit structure of Shigella cytotoxin." J Biol Chem **256**(16): 8732-8.
- Olsnes, S., C. Fernandez-Puentes, L. Carrasco and D. Vazquez (1975). "Ribosome inactivation by the toxic lectins abrin and ricin. Kinetics of the enzymic activity of the toxin A-chains." Eur J Biochem **60**(1): 281-288.
- Orita, M., F. Nishikawa, et al. (1996). "High-resolution NMR study of a GdAGA tetranucleotide loop that is an improved substrate for ricin, a cytotoxic plant protein." Nucleic Acids Res **24**(4): 611-8.
- Orita, M., F. Nishikawa, et al. (1993). "High-resolution NMR study of a synthetic oligoribonucleotide with a tetranucleotide GAGA loop that is a substrate for the cytotoxic protein, ricin." Nucleic Acids Res **21**(24): 5670-8.
- Patricia C. Weber, m. c. (1996). "structure-based design of thrombin inhibitors." 247-264.
- Pelham, H. R. (1990). "The retention signal for soluble proteins of the endoplasmic reticulum." Trends Biochem Sci **15**(12): 483-6.
- Pelham, H. R., L. M. Roberts, et al. (1992). "Toxin entry: how reversible is the secretory pathway?" Trends Cell Biol **2**(7): 183-5.
- Perez, C. and A. R. Ortiz (2001). "Evaluation of docking functions for protein-ligand docking." J Med Chem **44**(23): 3768-85.
- Perola, E., W. P. Walters, et al. (2004). "A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance." Proteins **56**(2): 235-49.
- Prestle, J., M. Schonfelder, et al. (1992). "Type 1 ribosome-inactivating proteins depurinate plant 25S rRNA without species specificity." Nucleic Acids Res **20**(12): 3179-82.
- Rapak, A., P. O. Falnes, et al. (1997). "Retrograde transport of mutant ricin to the endoplasmic reticulum with subsequent translocation to cytosol." Proc Natl Acad Sci U S A **94**(8): 3783-8.
- Rarey, M., B. Kramer, et al. (1996). "A fast flexible docking method using an incremental construction algorithm." J Mol Biol **261**(3): 470-89.
- Ready, M. P., B. J. Katzin, et al. (1988). "Ribosome-inhibiting proteins, retroviral reverse transcriptases, and RNase H share common structural elements." Proteins **3**(1): 53-9.

- Ready, M. P., Y. Kim, et al. (1991). "Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action." Proteins **10**(3): 270-8.
- Reisbig, R., S. Olsnes, et al. (1981). "The cytotoxic activity of Shigella toxin. Evidence for catalytic inactivation of the 60 S ribosomal subunit." J Biol Chem **256**(16): 8739-44.
- Richard A. Friesner, J. L. B., Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin (2004). "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy." journal of Medicinal chemistry **47**(7): 1739–1749.
- Richardson, P. T., M. Westby, et al. (1989). "Recombinant proricin binds galactose but does not depurinate 28 S ribosomal RNA." FEBS Lett **255**(1): 15-20.
- Riederer, M. A., T. Soldati, et al. (1994). "Lysosome biogenesis requires Rab9 function and receptor recycling from endosomes to the trans-Golgi network." J Cell Biol **125**(3): 573-82.
- Roberts, L. M. and J. M. Lord (1981). "The synthesis of Ricinus communis agglutinin, cotranslational and posttranslational modification of agglutinin polypeptides." Eur J Biochem **119**(1): 31-41.
- Roberts M. Stroud, J. F.-M. (2008). "computational and structural approaches to drug discovery."
- Robertus, J. (1991). "The structure and action of ricin, a cytotoxic N-glycosidase." Semin Cell Biol **2**(1): 23-30.
- Roday, S., T. Amukele, et al. (2004). "Inhibition of ricin A-chain with pyrrolidine mimics of the oxacarbenium ion transition state." Biochemistry **43**(17): 4923-33.
- Ruben Abagyan, M. T., D.A. Kuznetsov (1994). "ICM: a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation." Journal of Computational Chemistry **15**(5): 488-506.
- Rutenber, E., B. J. Katzin, et al. (1991). "Crystallographic refinement of ricin to 2.5 Å." Proteins **10**(3): 240-50.
- Rutenber, E. and J. D. Robertus (1991). "Structure of ricin B-chain at 2.5 Å resolution." Proteins **10**(3): 260-9.

- Sandvig, K., S. Grimmer, et al. (2002). "Pathways followed by ricin and Shiga toxin into cells." Histochem Cell Biol **117**(2): 131-41.
- Sandvig, K. and S. Olsnes (1980). "Diphtheria toxin entry into cells is facilitated by low pH." J Cell Biol **87**(3 Pt 1): 828-32.
- Sandvig, K., S. Olsnes, et al. (1979). "Inhibitory effect of ammonium chloride and chloroquine on the entry of the toxic lectin modeccin into HeLa cells." Biochem Biophys Res Commun **90**(2): 648-55.
- Sandvig, K., M. Ryd, et al. (1994). "Retrograde transport from the Golgi complex to the ER of both Shiga toxin and the nontoxic Shiga B-fragment is regulated by butyric acid and cAMP." J Cell Biol **126**(1): 53-64.
- Sandvig, K. and B. van Deurs (1996). "Endocytosis, intracellular transport, and cytotoxic action of Shiga toxin and ricin." Physiol Rev **76**(4): 949-66.
- Schlossman, D., D. Withers, et al. (1989). "Role of glutamic acid 177 of the ricin toxin A chain in enzymatic inactivation of ribosomes." Mol Cell Biol **9**(11): 5012-21.
- Schneider, G. and H. J. Bohm (2002). "Virtual screening and fast automated docking methods." Drug Discov Today **7**(1): 64-70.
- Schulz-Gasch, T. and M. Stahl (2003). "Binding site characteristics in structure-based virtual screening: evaluation of current docking tools." J Mol Model **9**(1): 47-57.
- Seidah, N. G., A. Donohue-Rolfe, et al. (1986). "Complete amino acid sequence of Shigella toxin B-chain. A novel polypeptide containing 69 amino acids and one disulfide bridge." J Biol Chem **261**(30): 13928-31.
- Seidler, J., S. L. McGovern, et al. (2003). "Identification and prediction of promiscuous aggregating inhibitors among known drugs." J Med Chem **46**(21): 4477-86.
- Shiga, K. (1898). "Ueber den Dysenterie-bacillus (Baccillus dysenteriae)." Zentralbl Bakteriol Orig. **24**: 913-918.
- Shoichet, B. K., S. L. McGovern, et al. (2002). "Lead discovery using molecular docking." Curr Opin Chem Biol **6**(4): 439-46.
- Simpson, J. C., C. Dascher, et al. (1995). "Ricin cytotoxicity is sensitive to recycling between the endoplasmic reticulum and the Golgi complex." J Biol Chem **270**(34): 20078-83.

- Simpson, J. C., L. M. Roberts, et al. (1999). "Ricin A chain utilises the endoplasmic reticulum-associated protein degradation pathway to enter the cytosol of yeast." FEBS Lett **459**(1): 80-4.
- Simpson, J. C., D. C. Smith, et al. (1998). "Expression of mutant dynamin protects cells against diphtheria toxin but not against ricin." Exp Cell Res **239**(2): 293-300.
- Stirpe, F. and M. G. Battelli (2006). "Ribosome-inactivating proteins: progress and problems." Cell Mol Life Sci **63**(16): 1850-66.
- Strockbine, N. A., M. P. Jackson, et al. (1988). "Cloning and sequencing of the genes for Shiga toxin from *Shigella dysenteriae* type 1." J Bacteriol **170**(3): 1116-22.
- Sturm, M. B., S. Roday, et al. (2007). "Circular DNA and DNA/RNA hybrid molecules as scaffolds for ricin inhibitor design." J Am Chem Soc **129**(17): 5544-50.
- Suh, J. K., C. J. Hovde, et al. (1998). "Shiga toxin attacks bacterial ribosomes as effectively as eucaryotic ribosomes." Biochemistry **37**(26): 9394-8.
- Szewczak, A. A. and P. B. Moore (1995). "The sarcin/ricin loop, a modular RNA." J Mol Biol **247**(1): 81-98.
- T.S. Ramalingam, P. K. D., Sunil K. Podder (1993). "identification of the adenine binding site in the ricin toxin a-chain by fluorescence, CD, and electron spin resonance spectroscopy." Biopolymers **33**: 1687-1694.
- Taylor, R. D., P. J. Jewsbury, et al. (2002). "A review of protein-small molecule docking methods." J Comput Aided Mol Des **16**(3): 151-66.
- Thaisrivongs, S., P. K. Tomich, et al. (1994). "Structure-based design of HIV protease inhibitors: 4-hydroxycoumarins and 4-hydroxy-2-pyrones as non-peptidic inhibitors." J Med Chem **37**(20): 3200-4.
- Thomke, s. (1998). "Modes of Experimentation: An Innovation Process - and Competitive - Variable." Research Policy (27): 315-332.
- Thompson, W. L., J. P. Scovill, et al. (1995). "Drugs that show protective effects from ricin toxicity in in vitro protein synthesis assays." Nat Toxins **3**(5): 369-77.
- Totrov, M. and R. Abagyan (1997). "Flexible protein-ligand docking by global energy optimization in internal coordinates." Proteins Suppl **1**: 215-20.

- Tully, R. E. and H. Beevers (1976). "Protein Bodies of Castor Bean Endosperm: Isolation, Fractionation, and the Characterization of Protein Components." Plant Physiol **58**(6): 710-716.
- Turnbull, W. B. and A. H. Daranas (2003). "On the value of c: can low affinity systems be studied by isothermal titration calorimetry?" J Am Chem Soc **125**(48): 14859-66.
- van Deurs, B., L. R. Pedersen, et al. (1985). "Receptor-mediated endocytosis of a ricin-colloidal gold conjugate in vero cells. Intracellular routing to vacuolar and tubulo-vesicular portions of the endosomal system." Exp Cell Res **159**(2): 287-304.
- van Deurs, B., T. I. Tonnessen, et al. (1986). "Routing of internalized ricin and ricin conjugates to the Golgi complex." J Cell Biol **102**(1): 37-47.
- Verdonk, M. L., J. C. Cole, et al. (2003). "Improved protein-ligand docking using GOLD." Proteins **52**(4): 609-23.
- von Itzstein, M., W. Y. Wu, et al. (1993). "Rational design of potent sialidase-based inhibitors of influenza virus replication." Nature **363**(6428): 418-23.
- Wallace, A. C., R. A. Laskowski, et al. (1995). "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions." Protein Eng **8**(2): 127-34.
- Wang, S., J. Feng, et al. (2005). "Structural-Based Rational Design of an Antagonist Peptide That Inhibits the Ribosome-Inactivating Activity of Ricin A Chain." Int. J. Pept. Res. Ther. **11**: 211-218.
- Warren, G. L., C. W. Andrews, et al. (2006). "A critical assessment of docking programs and scoring functions." J Med Chem **49**(20): 5912-31.
- Weber, P. C. and F. R. Salemme (2003). "Applications of calorimetric methods to drug discovery and the study of protein interactions." Curr Opin Struct Biol **13**(1): 115-21.
- Welch, W., J. Ruppert, et al. (1996). "Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites." Chem Biol **3**(6): 449-62.
- Wesche, J., A. Rapak, et al. (1999). "Dependence of ricin toxicity on translocation of the toxin A-chain from the endoplasmic reticulum to the cytosol." J Biol Chem **274**(48): 34443-9.
- White, J., L. Johannes, et al. (1999). "Rab6 coordinates a novel Golgi to ER retrograde transport pathway in live cells." J Cell Biol **147**(4): 743-60.

- Wiseman, T., S. Williston, et al. (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." Anal Biochem **179**(1): 131-7.
- Yan, X., P. Day, et al. (1998). "Recognition and interaction of small rings with the ricin A-chain binding site." Proteins **31**(1): 33-41.
- Yan, X., T. Hollis, et al. (1997). "Structure-based identification of a ricin inhibitor." J Mol Biol **266**(5): 1043-9.
- Youle, R. J. and A. H. Huang (1976). "Protein Bodies from the Endosperm of Castor Bean: Subfractionation, Protein Components, Lectins, and Changes during Germination." Plant Physiol **58**(6): 703-709.
- Zhang, J. H., T. D. Chung, et al. (1999). "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays." J Biomol Screen **4**(2): 67-73.
- Zsoldos, Z., D. Reid, et al. (2006). "eHiTS: an innovative approach to the docking and scoring function problems." Curr Protein Pept Sci **7**(5): 421-35.
- Zsoldos, Z., D. Reid, et al. (2007). "eHiTS: a new fast, exhaustive flexible ligand docking system." J Mol Graph Model **26**(1): 198-212.

Vita

Yan Bai was born in Shehezi, Xinjiang, P. R. China, on August 4th, 1977, the daughter of Yongzhi Bai and Xiuzhi Zhang. She graduated from Ocean University of China in 2000 with a Bachelor of Medicinal Chemistry. She began her graduate study in Plant Biology in 2003 September at The University of Texas at Austin, and then transferred to Chemistry and Biochemistry department. Under the supervision of Dr. Jon D Robertus, she has completed doctoral training in the Department of Biochemistry. In December of 2009, Yan graduated from The University of Texas at Austin with the degree of Doctor of Philosophy.

Permanent address: Cuiyuan Xincun Building #10, Room #412, Shihezi, Xinjiang, P.R.
China 832000

This dissertation was typed by the author.