Published in final edited form as:

Methods Mol Biol. 2013 ; 993: 177–184. doi:10.1007/978-1-62703-342-8_12.

Identification of Novel Anthrax Toxin Countermeasures Using In Silico Methods

Ting-Lan Chiu, Kimberly M. Maize, and Elizabeth A. Amin

Abstract

Anthrax is an acute infectious disease caused by the spore-forming, gram-positive, rod-shaped bacterium *Bacillus anthracis*. The anthrax toxin lethal factor (LF) is the primary anthrax toxin component responsible for cytotoxicity and host death and has been a heavily researched target for design of postexposure therapeutics in the event of a bioterror attack. Various computer-aided drug design methodologies have proven useful for pinpointing new antianthrax drug scaffolds, optimizing existing leads and probes, and elucidating key mechanisms of action. We present a selection of in silico virtual screening protocols incorporating docking and scoring, shape-based searching, and pharmacophore mapping techniques to identify and prioritize small molecules with potential biological activity against LF. We also recommend screening parameters that have been shown to increase the accuracy and reliability of these computational results.

Keywords

Anthrax; Lethal factor; Metalloproteinases; Virtual screening; Docking and scoring; Pharmacophore mapping; Computer-aided drug design

1 Introduction

Molecular modeling techniques, including docking and scoring, shape-based searching, and pharmacophore mapping, are widely used in the drug discovery process to identify new molecular scaffolds, elucidate mechanisms of action, and prioritize particular compounds and/or series for experimental screening or synthesis (1–5). Incorporating virtual screening (VS) methodologies such as docking and topomeric searching into therapeutics design has had a positive impact on compound hit rates (1) and has led to better prediction of binding modes (1, 3, 4). However, the reliability of VS simulations varies broadly among the available docking algorithms, scoring functions, parameters, and descriptors (Maize and Amin, unpublished observation, 2012, (6–8)), which must be chosen carefully on the basis of ligand structure(s) and key receptor characteristics such as hydrophobicity and hydrogen-bonding environment. Docking and scoring in particular are often conducted using default settings and parameters available in the VS package on hand; however, variations in binding-site structure and features, as well as applied tolerance thresholds for steric and electrostatic interactions, strongly influence the quality of results. Validation studies should be conducted whenever possible to assess the capability of VS protocols to accurately reproduce experimental bound configurations to the target of interest and to prioritize known

active compounds within a dataset. Metalloenzymes, such as the anthrax toxin lethal factor (LF), which is primarily responsible for anthrax-related toxicity (9–13), have attracted particular attention as drug targets in various disease modalities. However, they pose significant challenges in VS (and in other molecular modeling techniques as well) owing to the presence of catalytic transition metals that are often poorly represented in molecular mechanics-based force fields used to construct scoring functions. We outline a series of experimentally validated protocols and recommended parameter sets for docking and scoring, shape-based (topomeric) searching, and pharmacophore perception that have been shown to increase experimental hit rates for small molecules targeting LF and that can be implemented to rank existing compounds for further prosecution and to search databases for new LF inhibitor scaffolds.

2 Materials

1. SciTegic Pipeline Pilot 8.0, Accelrys, Inc. (San Diego, CA)
2. SYBYL-X 1.0 expert molecular modeling environment, Tripos, Inc. (St. Louis, MO)
3. Surflex-Dock 2.1 virtual screening package (14–17) and CScore consensus scoring module (18), Tripos, Inc.
4. Topomer Search (19–22) three-dimensional (3D) shape-based searching module, Tripos, Inc.
5. MOE (Molecular Operating Environment) 2010.10, Chemical Computing Group, Inc. (Montreal, Quebec, Canada)
6. GALAHAD (Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Datasets) (23) pharmacophore hypothesis module, Tripos, Inc.

3 Methods

1. Small molecules of interest can be prepared for LF screening via a variety of mechanisms. For a relatively small library, compounds can be sketched individually and then geometry optimized in SYBYL-X or MOE using an appropriate force field for drug-like small molecules, such as Tripos (24) or MMFF94s (Merck Molecular Force Field) (25), in order to obtain 3D coordinates for subsequent simulations. For a wider-ranging virtual screen (e.g., to identify new scaffolds in less explored areas of chemistry space), small-molecule compound libraries can be obtained from a variety of sources, including DrugBank (26, 27) (<http://www.drugbank.ca/>), the National Institutes of Health (NIH) Molecular Libraries Small Molecule Repository (MLSMR) (28), the National Cancer Institute (29), and eMolecules (<http://www.emolecules.com/>). For these larger compound sets, we recommend generating 3D configurations using SciTegic Pipeline Pilot via the “SD Reader,” “3D Coordinates,” “Add Hydrogens,” “Minimize Molecule,” and “SD Writer” components, in that order. This protocol breaks each compound into ring and chain fragments, generates 3D structures for the fragments, reassembles the compound, conducts a preliminary geometry optimization on the reassembled

structure, adds hydrogen atoms, carries out a more thorough energy minimization via the Clean force field (30), and writes a new 3D SD file as output that can subsequently be used as input for a variety of software packages and modeling techniques.

2. For receptor-based procedures including docking and scoring, five experimental X-ray structures of LF-ligand complexes are currently available in the Protein Data Bank (<http://www.rcsb.org/pdb/download/download.do>): 1YQY (31), 1ZXV (32), 1PWP (33), 1PWQ (34), and 1PWU (34). Cocrystallized inhibitors in these complexes include a sulfonamide hydroxamate, MK-702/LF-1B, the most active LF inhibitor designed to date (half-maximal inhibitory concentration $[IC_{50}] = 0.054 \mu\text{M}$, 1YQY), rhodanine derivative BI-MFM3 ($IC_{50} = 1.7 \mu\text{M}$, 1ZXV), the *N*, *N'*-di-quinoline urea analogue NSC 12155 ($K_i = 0.5 \mu\text{M}$, 1PWP), and two peptidic hydroxamates, thioacetyl-Tyr-Pro-Met amide and GM6001 ($K_i^{\text{app}} = 2.1 \mu\text{M}$ and $11 \mu\text{M}$, 1PWQ and 1PWU, respectively). We recommend 1YQY.pdb (31) for VS, because it is a truncated LF structure comprising the three key domains (II–IV) that form the enzyme active site. Protein and ligand preparation can be done in MOE 2010.10: remove all cocrystallized water molecules; add hydrogens, check ligand bonds/protonation state(s) and edit if necessary; examine residue bonds/protonation states within $4.5 \text{ \AA}$ of the cocrystallized ligand and correct if necessary; fix heavy atoms in space and then energy minimize the complex using an initial gradient of 0.05.
3. Docking and scoring can be carried out using Surflex-Dock (14–17) and CScore (18) in SYBYL-X 1.0. In detailed evaluations (1) of various docking and scoring programs for the LF system, Surflex-Dock was found to be superior in terms of reproducing cocrystallized LF-inhibitor bound configurations, within root-mean-square-deviation (RMSD) values of only $0.54 \text{ \AA}$. In this VS environment, the target area on the receptor for small-molecule docking is rendered by a protomol, which is a representation of an “ideal” ligand located in the active site of the protein, constructed by modeling specific interactions of small-molecule, fragment-based probes within that area (14–17). Probes include a variety of hydrophobic and hydrophilic fragments as well as hydrogen-bond donor and acceptor groupings. In the screening procedure, the targeted area in the active site can be obtained from an automatic detection algorithm, the location of the cocrystallized ligand, or manual selection of relevant residues using the SYBYL interface. Extensive validation studies (Maize and Amin, unpublished observation, 2012) on experimental LF X-ray structures and their cocrystallized ligands indicated that ligand-based protomol generation *without* hydrogens added to the ligand yielded greater screening accuracy, in terms of RMSD between predicted and experimental bound configurations as well as docking enrichment factor (35), compared with automatic site detection or manual residue selection.
4. Choosing acceptable threshold and bloat values is critical to maximizing LF docking precision. In Surflex-Dock, the threshold value determines how much of the protomol may be “buried” within the protein, while the bloat parameter allows

for protomol expansion in order to reach into crevices or longer binding channels (especially those with open ends). Evaluating a variety of threshold values from 0.01 to 1 for docking the five available cocrystallized LF inhibitors into their respective crystal structures (31–34) led to an optimal threshold value of 0.74 and a bloat value of zero (Amin and Maize, unpublished observation, 2012). These studies also pinpointed two key user-defined variables that impact LF docking outcomes: ring flexibility and search density (14–16). Including ring flexibility (a binary operator) and implementing the highest available density of search level ($d = 9$) increase computation time but significantly improve the quality of results and are therefore recommended for all LF VS runs. The maximum number of conformations per compound fragment and the maximum number of poses per ligand should be set to 20, with the maximum number of rotatable bonds per small molecule set to 100. Postdock minimizations are recommended for each bound configuration and, given the significant amount of steric and electrostatic variation in the LF active site, all four consensus scoring functions in CScore (G_SCORE, PMF_SCORE, D_SCORE, and CHEMSCORE) should be implemented in order to represent the broadest possible selection of ligand–receptor interactions taking place in the sterically and electrostatically diverse LF binding site.

5. Shape-based, “topomeric” searching can be done in order to find new LF inhibitor scaffolds occupying un- or underexplored regions in chemical space. In this procedure, one or more proven active compounds are utilized to search collections of molecules for matches that exhibit similar 3D shapes, as represented by conformationally independent topomeric fields (19–22). We have found this similarity searching protocol to be highly useful for identifying active LF inhibitors within data collections of various sizes and diversity levels as well as for selecting compounds for subsequent experimental evaluation via in vitro screening (1). Structures pinpointed by topomeric searching are often significantly dissimilar in terms of usual two-dimensional (2D) structural fingerprints, meaning that they are more likely to be located in less extensively explored chemical space than those identified by traditional 2D similarity searching (19–22). One often uses a highly active but pharmacokinetically “compromised” compound as the topomeric search template to “lead-hop” to new compounds that have similar 3D shapes (and, ostensibly, similar ability to bind to the desired target) but different chemical functionalities, in order to retain biological activity while avoiding impediments such as toxicity and metabolic instability. Effective searching using an active LF inhibitor template can be done using the Topomer Search module in SYBYL-X 1.0, using a “maximum distance considered hit” parameter of 185, with all weighting factors (steric, aromatic, positive/negative, donor/acceptor) set to 1,000.
6. Accurate and validated pharmacophore hypotheses have proven useful for identifying new LF inhibitor scaffolds via database searching on the basis of ligand–receptor interactions observed for one or more series of active compounds (33, 36–38). Several LF inhibitor pharmacophore hypotheses have been outlined in the literature (33, 36–38) ; however, these models were developed from relatively small training sets that occupy only one or two subsites of the LF active site and

therefore do not necessarily represent the majority of key interactions that are essential for ligand binding. We recently reported (2) a new comprehensive pharmacophore map based on experimentally determined bound configurations for active compounds; this new hypothesis covers all three subsites (S1', S1–S2, and S2') of the LF active site and selectively identifies inhibitors with biological activity against LF in the nanomolar range. As reported by Chiu and Amin (2), for accurate and useful pharmacophore mapping based on active LF inhibitors, we recommend a genetic algorithm approach incorporating Pareto scoring, as implemented in the GALAHAD pharmacophore perception module (23) (see Note 1), together with ligand–receptor interaction analysis based on experimental structural biology (see Note 2).

7. If experimental bound configurations (i.e., cocrystallized inhibitors) are to be used for pharmacophore perception, we recommend aligning all structures in Cartesian space by optimizing the sum of all pairwise alignment scores using the Homology/Align module in MOE 2010.10, basing the alignment on protein coordinates. Structures of additional small molecules can be subjected to geometry optimization by energy minimization within the LF X-ray structure of choice (we suggest 1YQY.pdb) in order to approach a putative bound configuration as closely as possible. Optimization can be done in MOE 2010.10 using the MMFF94s force field (25) with a convergence criterion of 0.05 kcal/mol · Å, with the receptor held rigid. Larger sets of molecules used for hypothesis validation and/or database searching can be prepared and optimized using MMFF94s and then docked into the LF active site using Surflex-Dock and CScore as described above, with the protomol defined to encompass all three LF binding area subsites and threshold and bloat parameters set to the optimal values of 0.74 and zero.

Acknowledgments

This work was supported by NIH R01 AI083234 to E.A.A.; the Minnesota Supercomputing Institute for Advanced Computational Research (MSI); and the University of Minnesota Institute for Therapeutics Discovery and Development (ITDD).

¹All genetic algorithm–based pharmacophore hypotheses should be subjected to multiple iterative refinement, with accuracy assessed by two criteria: (1) an overall Pareto score and (2) a rank sum value that includes the GALAHAD parameters of steric overlap, pharmacophoric concordance, and agreement between the query tuple and the pharmacophoric tuples for the ligands used to create the model (which is essentially a similarity value between the query and the ligand set) (23). Generally, if all Pareto scores are equal, the models are ordered by the rank sum value, with any remaining “ties” broken by a total strain energy term where lower energy is considered more favorable. Recommended user-specified parameters for LF inhibitors (to be fine-tuned based on the training set used) include a population size of 25–35; a maximum number of 90 generations; three to five molecules that must hit the query in order for a model to be retained; and a “keep best *n* models” value of 15–20.

²The presence of a given chemical functionality in more than one compound used to generate a pharmacophore hypothesis does not guarantee a significant contribution to activity. It is therefore helpful to examine structural biology data for ligand–receptor complexes whenever possible and to model experimentally observed interactions as 2D ligand–receptor interaction maps (in MOE 2010.10). Pharmacophoric features that either do not parallel experimental interactions or represent those interactions inaccurately (e.g., incorrect hydrogen-bonding directionality) can then be removed from the hypotheses of interest. Because hydrophobic interactions are not well rendered in MOE 2D interaction maps, supplemental PoseView (<http://www.zbh.uni-hamburg.de/poseview>) (39) 2D diagrams can be generated in cases such as LF where hydrophobic interactions demonstrate a significant contribution to compound activity.

References

1. Chiu T, Solberg J, Patil S, et al. Identification of novel non-hydroxamate anthrax toxin lethal factor inhibitors by topomeric searching, docking and scoring, and in vitro screening. *J Chem Inf Model*. 2009; 49:2726–2734. [PubMed: 19928768]
2. Chiu TL, Amin EA. Development of a comprehensive, validated pharmacophore hypothesis for anthrax toxin lethal factor (LF) inhibitors using genetic algorithms, Pareto scoring, and structural biology. *J Chem Inf Model*. 2012; 52:1886–1897. [PubMed: 22697455]
3. Amin EA, Welsh WJ. Highly predictive CoMFA and CoMSIA models for two series of stromelysin-1 (MMP-3) inhibitors elucidate S1' and S1–S2' binding modes. *J Chem Inf Model*. 2006; 46:1775–1783. [PubMed: 16859309]
4. Jia Y, Chiu TL, Amin EA, et al. Design, synthesis and evaluation of analogs of initiation factor 4E (eIF4E) cap-binding antagonist Bn7-GMP. *Eur J Med Chem*. 2010; 45:1304–1313. [PubMed: 20060622]
5. Ambrose Amin E, Welsh WJ. Three-dimensional quantitative structure-activity relationship (3D-QSAR) models for a novel class of piperazine-based stromelysin-1 (MMP-3), inhibitors: applying a “divide and conquer” strategy. *J Med Chem*. 2001; 44:3849–3855. [PubMed: 11689071]
6. Moustakas DT, Lang PT, Pegg S, et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des*. 2006; 20:601–619. [PubMed: 17149653]
7. Hartshorn MJ, Verdonk ML, Chessari G, et al. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem*. 2007; 50:726–741. [PubMed: 17300160]
8. Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model*. 2007; 47:435–449. [PubMed: 17305329]
9. Pezard C, Berche P, Mock M. Contribution of individual toxin components to virulence of *Bacillus anthracis*. *Infect Immun*. 1991; 59:3472–3477. [PubMed: 1910002]
10. Chopra AP, Boone S, Liang X, et al. Anthrax lethal factor proteolysis and inactivation of MAPK kinase. *J Biol Chem*. 2003; 278:9402–9406. [PubMed: 12522135]
11. Vitale G, Bernardi L, Napolitani G, et al. Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochem J*. 2000; 352:739–745. [PubMed: 11104681]
12. Moayeri M, Leppla SH. The roles of anthrax toxin in pathogenesis. *Curr Opin Microbiol*. 2004; 7:19–24. [PubMed: 15036135]
13. Warfel JM, Steele AD, D'Agnillo F. Anthrax lethal toxin induces endothelial barrier dysfunction. *Am J Pathol*. 2005; 166:1871–1881. [PubMed: 15920171]
14. Jain AN. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des*. 2007; 21:281–306. [PubMed: 17387436]
15. Pham T, Jain AN. Parameter estimation for scoring protein-ligand interactions using negative training data. *J Med Chem*. 2006; 49:5856–5868. [PubMed: 17004701]
16. Jain AN. Virtual screening in lead discovery and optimization. *Curr Opin Drug Discov Devel*. 2004; 7:396–403.
17. Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity- based search engine. *J Med Chem*. 2003; 46:499–511. [PubMed: 12570372]
18. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. *J Comput Chem*. 1992; 13:505–524.
19. Cramer RD, Poss MA, Hermsmeier MA, et al. Prospective identification of biologically active structures by topomer shape similarity searching. *J Med Chem*. 1999; 42:3919–3933. [PubMed: 10508440]
20. Cramer RD, Jilek RJ, Guessregen S, et al. Lead hopping. Validation of topomer similarity as a superior predictor of similar biological activities. *J Med Chem*. 2004; 47:6777–6791. [PubMed: 15615527]
21. Cramer RD. Leadhopping – and beyond. *Expert Opin Drug Discov*. 2006; 1:311–321. [PubMed: 23495902]

22. Jilek RJ, Cramer RD. Topomers: a validated protocol for their self-consistent generation. *J Chem Inf Comput Sci*. 2004; 44:1221–1227. [PubMed: 15272829]
23. Richmond NJ, Abrams CA, Wolohan PR, et al. GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. *J Comput Aided Mol Des*. 2006; 20:567–587. [PubMed: 17051338]
24. Clark M, Cramer RD III, Van Opdenbosch N. Validation of the general purpose tripos 5.2 forcefield. *J Comput Chem*. 1989; 10:982–1012.
25. Halgren TA. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J Comput Chem*. 1999; 20:730–748.
26. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008; 36:D901–D906. [PubMed: 18048412]
27. Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006; 34:D668–D672. [PubMed: 16381955]
28. Austin CP, Brady LS, Insel TR, et al. NIH molecular libraries initiative. *Science*. 2004; 306:1138–1139. [PubMed: 15542455]
29. [Accessed 29 Mar 2012] National Cancer Institute. [updated 16 Mar 2010]. <http://cactus.nci.nih.gov/download/nci/>.
30. Hahn M. Receptor surface models. 1 Definition and construction. *J Med Chem*. 1995; 38:2080–2090. [PubMed: 7783139]
31. Shoop WL, Xiong Y, Wiltse J, et al. Anthrax lethal factor inhibition. *Proc Natl Acad Sci USA*. 2005; 102:7958–7963. [PubMed: 15911756]
32. Forino M, Johnson S, Wong TY, et al. Efficient synthetic inhibitors of anthrax lethal factor. *Proc Natl Acad Sci USA*. 2005; 102:9499–9504. [PubMed: 15983377]
33. Panchal RG, Hermone AR, Nguyen TL, et al. Identification of small molecule inhibitors of anthrax lethal factor. *Nat Struct Mol Biol*. 2004; 11:67–72. [PubMed: 14718925]
34. Turk BE, Wong TY, Schwarzenbacher R, et al. The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nat Struct Mol Biol*. 2004; 11:60–66. [PubMed: 14718924]
35. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem*. 2006; 49:6789–6801. [PubMed: 17154509]
36. Yuan H, Johnson SL, Chen LH, et al. A novel pharmacophore model for the design of anthrax lethal factor inhibitors. *Chem Biol Drug Des*. 2010; 76:263–268. [PubMed: 20572812]
37. Agrawal A, de Oliveira CA, Cheng Y, et al. Thioamide hydroxypyrothiones supersede amide hydroxypyrothiones in potency against anthrax lethal factor. *J Med Chem*. 2009; 52:1063–1074. [PubMed: 19170530]
38. Roy J, Kumar UC, Machiraju PK, et al. In silico studies on anthrax lethal factor inhibitors: pharmacophore modeling and virtual screening approaches towards designing of novel inhibitors for a killer. *J Mol Graph Model*. 2010; 29:256–265. [PubMed: 20727800]
39. Stierand K, Maass PC, Rarey M. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics*. 2006; 22:1710–1716. [PubMed: 16632493]