



Research Techniques Made Simple: Molecular Docking in Dermatology - A Foray into In Silico Drug Discovery

Naiem T. Issa¹, Evangelos V. Badiavas¹ and Stephan Schürer²

Drug discovery is a complex process with many potential pitfalls. To go to market, a drug must undergo extensive preclinical optimization followed by clinical trials to establish its efficacy and minimize toxicity and adverse events. The process can take 10–15 years and command vast research and development resources costing over \$1 billion. The success rates for new drug approvals in the United States are < 15%, and investment costs often cannot be recouped. With the increasing availability of large public datasets (big data) and computational capabilities, data science is quickly becoming a key component of the drug discovery pipeline. One such computational method, large-scale molecular modeling, is critical in the preclinical hit and lead identification process. Molecular modeling involves the study of the chemical structure of a drug and how it interacts with a potential disease-relevant target, as well as predicting its ADMET properties. The scope of molecular modeling is wide and complex. Here we specifically discuss *docking*, a tool commonly employed for studying drug-target interactions. Docking allows for the systematic exploration of how a drug interacts at a protein binding site and allows for the rank-ordering of drug libraries for prioritization in subsequent studies. This process can be efficiently used to virtually screen libraries containing over millions of compounds.

Journal of Investigative Dermatology (2019) **139**, 2400–2408; doi:10.1016/j.jid.2019.06.129

CME Activity Dates: 18 November 2019

Expiration Date: 17 November 2020

Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: Evangelos V. Badiavas is the founder of Aegle Therapeutics. All other authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationships with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:

- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of Beaumont Health and the Society for Investigative Dermatology. Beaumont Health is accredited by the ACCME to provide continuing medical education for physicians. Beaumont Health designates this enduring material for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*[™]. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the *Journal of Investigative Dermatology* website: <http://www.jidonline.org/current>. Tests for CME credits may only be submitted online at <https://beaumont.cloud-cme.com/RTMS-Dec19> — click ‘CME on Demand’ and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

INTRODUCTION

Molecular recognition is a critical event in drug-protein interactions. Multiple theories regarding how a protein recognizes a drug exist, such as the lock and key, induced fit, and

conformational selection models. Generally, a drug enters the protein binding pocket and interacts with the amino acid (residue) side chains in the pocket to form a complex—a process typically governed by non-covalent bonding

¹Dr. Phillip Frost Department of Dermatology and Cutaneous Surgery, University of Miami School of Medicine, Miami, Florida, USA; and ²Department of Molecular and Cellular Pharmacology, University of Miami School of Medicine, Miami, Florida, USA

Correspondence: Naiem T. Issa, Department of Dermatology and Cutaneous Surgery, Department of Molecular and Cellular Pharmacology, University of Miami School of Medicine, Miami, Florida 33136. E-mail: nti4@med.miami.edu

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; ADMET, absorption; distribution, metabolism; excretion, toxicity

SUMMARY POINTS

- Computational molecular modeling tools aid in drug discovery to increase throughput and accuracy in preclinical lead identification and optimization
- Docking is the computational modeling of how drugs can occupy and interact with protein target binding sites
- Docking aids in narrowing the potential chemical (drug) space from millions of compounds to tens or hundreds for efficient biological testing and validation
- Open source software for molecular modeling and docking are tools developed in academic settings freely available for academicians to use in their investigations

interactions (Kitchen et al., 2004). These interactions include hydrogen bonding, van der Waals forces, metal coordination, hydrophobic forces, pi-pi interactions, and electrostatic interactions. Thermodynamically, this protein-ligand binding can be quantified by the free energy of binding (ΔG), which is the free energy of the protein-ligand complex minus the free energy of the protein and the ligand in their unbound states. The greater (more negative) the ΔG , the greater the stability of the ensuing complex and the more likely its formation (Figure 1). ΔG includes enthalpy and entropy ($\Delta G = \Delta H - T\Delta S$) and is directly related to the binding constant ($\Delta G = RT \ln K_d$). It is important to recognize that binding free energy, in addition to non-binding interactions mentioned above include solvation and desolvation, as well as the internal energies of the protein and ligand (e.g., owing to conformational changes upon binding and strain).

Molecular modeling aims to study drug-protein recognition through the computational calculation of physical forces. Specifically, molecular docking is a widely used technique to systematically explore how a drug interacts in a protein binding site, considering its conformations and orientations, or poses, and the energetics of the interactions in the protein-ligand complex to estimate the relative binding affinity (ΔG) in a computationally efficient way via a so called scoring function (Figure 2). In this sense, docking can be used to predict the potential binding pose of a drug and the most likely protein residue interactions based on the lowest (estimated relative) ΔG (Kitchen et al., 2004). Docking scoring functions typically combine empirical terms (parametrized based on known binding affinities) and energy calculations based on the specific types of interactions (e.g., electrostatic, hydrophobic). Entropic contributions, such as ligand rotatable bond restriction, may also be approximated depending on the scoring function used. A docking score can, thus, be considered an estimated relative ΔG ; it is typically a relative score useful to rank-order drugs and poses with respect to a protein target but not an exact globally comparable binding ΔG . Accuracy is also dependent on the scoring function used,

as each has been trained from different datasets (Wang et al., 2016). Nonetheless, a great utility of docking is to prioritize a set of compounds for biological testing from extremely large drug databases (e.g., over 100 million compounds) against one or more proteins. As the biological high-throughput screening of millions of compounds against a target is generally economically or technically unfeasible, docking allows investigators to expend resources on testing compounds with the greatest likelihood of interacting with their target. It is now computationally feasible to dock hundreds of millions of compounds and to identify novel activity from only a small number (< 1,000) of the best ranked compounds (Lyu et al., 2019). In contrast to biological screening, docking, or computational screening in general, does not require a physical sample. It is therefore possible to explore a very large chemical space of up to hundreds of millions of compounds and then acquire only the most likely active compounds for biological testing.

In addition, docking can be used for understanding the specific binding mode of drugs previously determined to bind a target. This information is highly valuable, as further medicinal chemistry optimizations may be pursued to improve binding against the target, avoid binding to undesirable off-targets, and also optimize ADMET.

DOCKING IN ACTION: EXAMPLES OF USE IN DRUG DISCOVERY

One of the first implementations of docking was by Desjarlais and Dixon (Desjarlais and Dixon, 1994) for HIV drug discovery. They docked known HIV protease inhibitors into its binding site and consequently designed a more potent inhibitor with an inhibitory constant (K_i) of 48 μM (Chenera et al., 1993). Zhao et al. utilized docking to discover novel inhibitors of JAK2, a critical member of the JAK-STAT signal transduction pathway (Zhao et al., 2015). Using the AZD1480 inhibitor-bound JAK2 crystal structure as a reference, they docked 3,010 drug-like molecules into the ATP-binding pocket and tested the top 10 compounds exhibiting the best scores in an inhibitory assay. They found the aminopyridine ethyl 1-(5-([3-methoxyphenyl]carbamoyl)-3-nitropyridin-2-yl)piperidine-4-carboxylate to be the most potent inhibitor, and subsequent optimization resulted in a low-micromolar inhibitory profile. Later studies used docking to screen large compound libraries against their protein of interest. For example, Mirza et al. used docking to screen 18 million compounds against dengue virus nonstructural protein 3 (NS3) (Mirza et al., 2016), and five inhibitors were identified with the ability to reduce virus titers in HUH7 cells (Mirza et al., 2018).

In the context of dermatology, Mann et al. (Mann et al., 2018) studied why most inhibitors against tyrosinase, the rate-limiting enzyme of melanin production, lack clinical efficacy. Inhibitors have traditionally been identified using mushroom tyrosinase. The authors screened a 50,000-compound library against recombinant human tyrosinase (hTyr) in vitro and found Thiamidol to have strong inhibition compared with hydroxyquinone and kojic acid but weak inhibition against mushroom tyrosinase. The docking of Thiamidol to hTyr revealed that it interacts with hydrophobic amino acids that are not found in mushroom tyrosinase

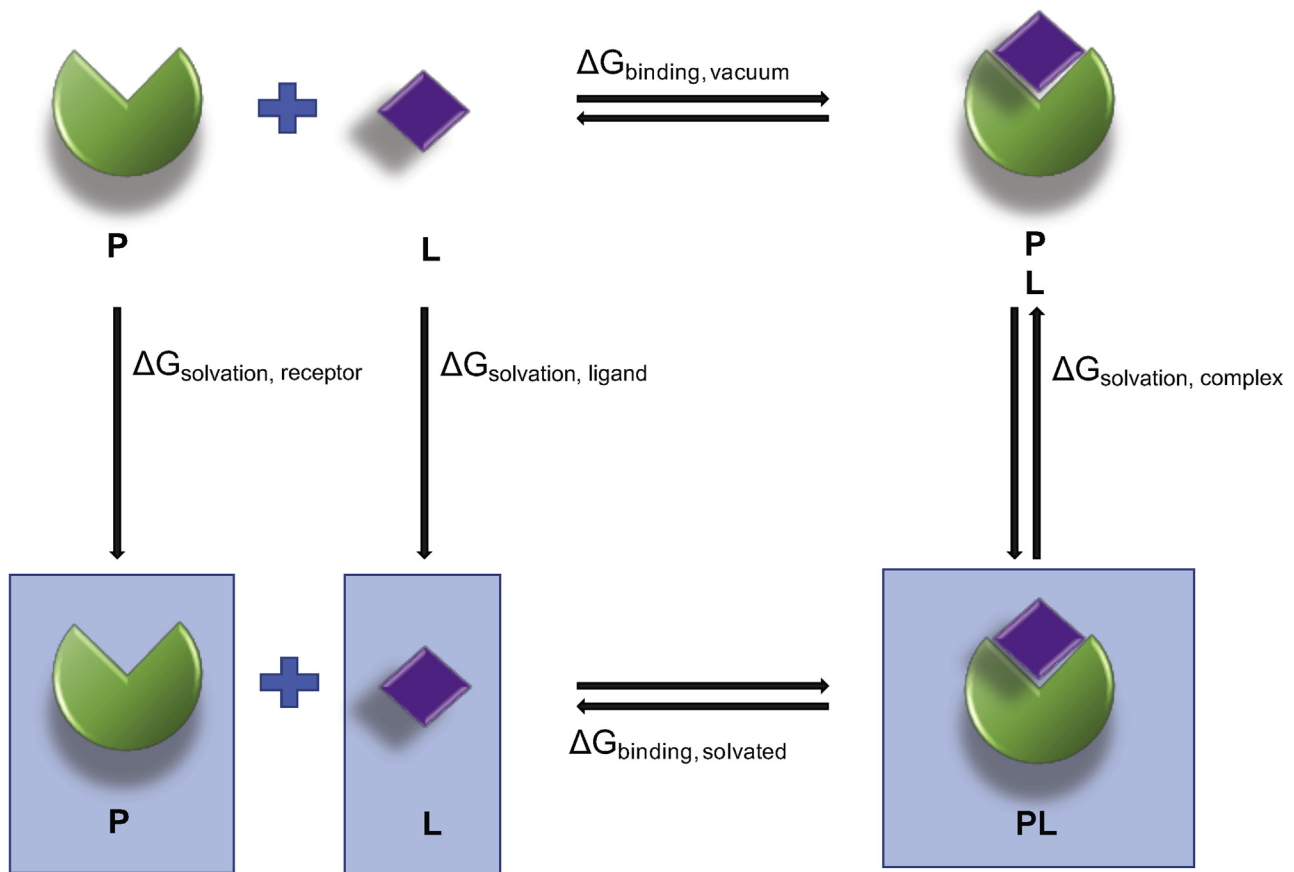


Figure 1. Simplified schematic of the protein (P) – ligand (L) binding process. P-L binding or dissociation are governed by the free energy of binding ($\Delta G_{\text{binding}}$), also known as binding affinity. Physiologically, binding events occur in an aqueous solvent (e.g., cellular or interstitial fluid). Ligand binding causes water to be displaced from the protein binding site and from around the ligand, a processes termed desolvation. Thus, energy calculations must consider the solvation energy, which is the energy needed to move the protein, ligand, or P-L complex from a vacuum to the solvent (termed “solvation”). Ultimately, the $\Delta G_{\text{binding}}$ is determined by complex interaction energy ($\Delta G_{\text{complex}}$) subtracted by the protein and ligand solvation energies ($\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - \Delta G_{\text{solvation, receptor}} - \Delta G_{\text{solvation, ligand}}$).

(Figure 3), possibly explaining the differential effect. Subsequent clinical testing showed the efficacy of Thiamidol in reducing the appearance of age spots.

Recently, Ghosh et al. (Ghosh et al., 2018) employed docking to identify a novel treatment for acne. The authors first designed a library of molecules with a quinolone backbone, known to have anti-bacterial properties by binding to bacterial DNA gyrase, and a nitro-heterocyclic motif, also known to have anti-bacterial and anti-inflammatory features, arranged in different spatial orientations. These molecules were subjected to docking to ascertain their potential to inhibit DNA gyrase. Prior to docking, they utilized the protein crystal structure of *Staphylococcus aureus*, since no crystal structure of *Propionibacterium acnes* DNA gyrase exists. The structure of *S. aureus* DNA gyrase was used, as it had > 40% amino acid sequence identity with that of *P. acnes*. VCD-004 was predicted to bind more optimally to DNA gyrase and interact with amino acids that are different from those that clindamycin and nadifloxacin (positive controls) interact with and become mutated to confer drug resistance in *P. acnes*. Therefore, VCD-004 can escape known methods of drug resistance. In vitro and in vivo testing confirmed the in silico studies. The discovery of VCD-004 as a novel antibiotic

against *P. acnes* for the treatment of drug-resistant acne demonstrates how docking and molecular modeling are implemented in rational drug design.

Drug discovery in dermatology also faces unique challenges regarding pharmacokinetics. The multiple layers of skin are a barrier against drug diffusion, and skin cells such as keratinocytes contain drug-metabolizing enzymes that could affect bioavailability (van Eijl et al., 2012). Adverse effects may also be because of metabolic products of the parent drug. As such, docking can help identify the likely metabolizing enzyme of a drug and predict what metabolite(s) may form (Sevrioukova and Poulos, 2015). The metabolite(s) can further be docked to target proteins to predict whether they will have additional effects beyond that of the parent drug.

A GENERALIZED PROTOCOL FOR USING MOLECULAR DOCKING

For investigators interested in applying docking to their studies, we point the reader to the article by Forli et al. (Forli et al., 2016) for a detailed protocol using the open source AutoDock platform (Trott and Olson, 2010). We present here a generalized protocol applicable to investigators of any skill level (Figure 4). However, this protocol is an

RESEARCH TECHNIQUES MADE SIMPLE

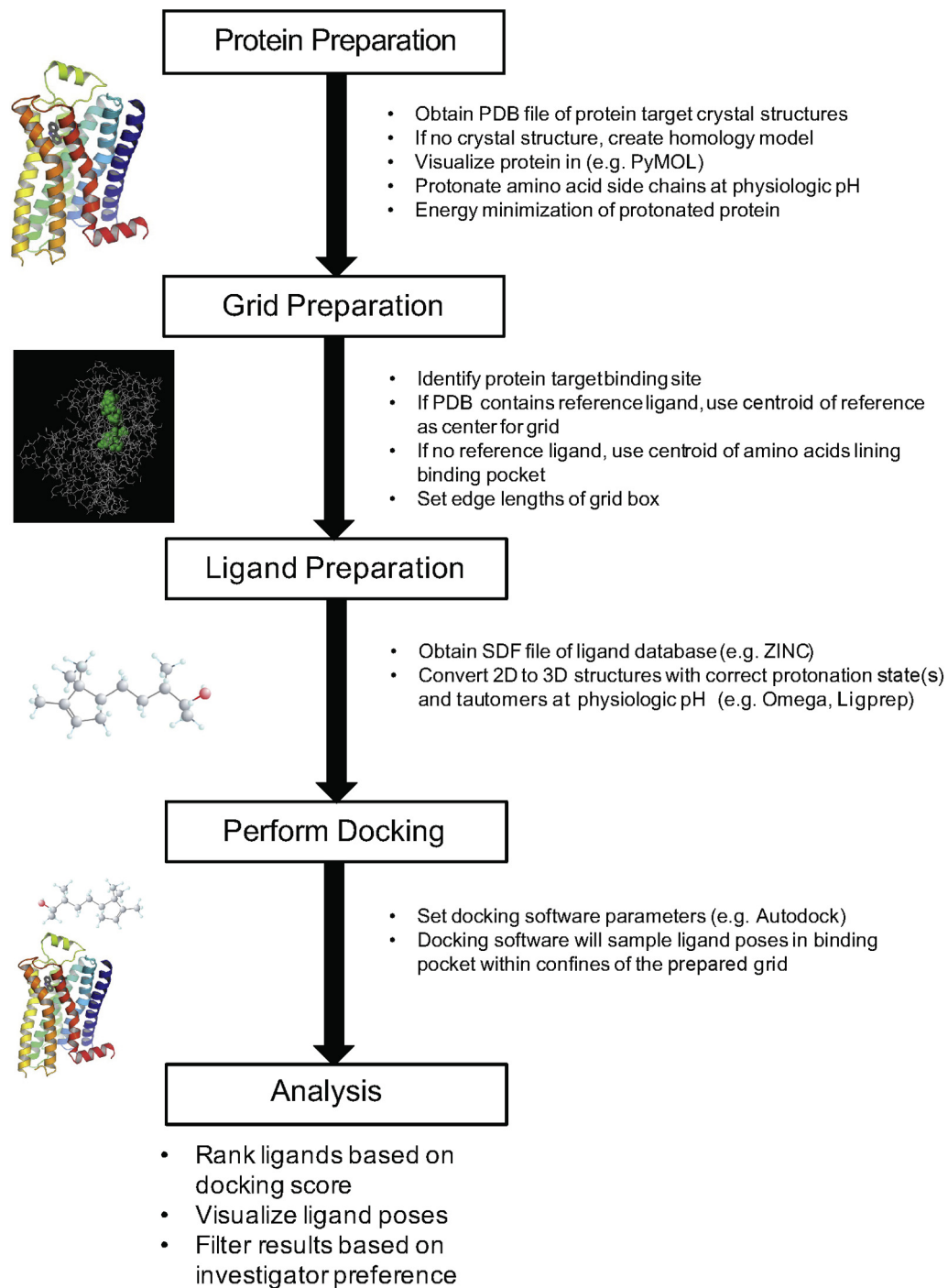


Figure 4. Schematic of a generic docking workflow. A detailed protocol can be found in (Forli et al., 2016).

contains information on the atoms of the protein, where they are in space (three-dimensional Cartesian coordinates), residues they are associated with, and how they are connected (e.g., secondary structure). Sometimes the protein is co-crystallized with other molecules, such as formaldehyde solvent or bound ligands. These molecules are also represented in the PDB file and may need to be removed or manipulated. Once the PDB file is obtained, the investigator may study the structure using visualization software such as VMD (Humphrey et al., 1996). The protein will need to be

further prepared by assigning bond orders, treating non-standard atoms, and assigning the appropriate protonation states on functional groups (e.g., aspartic acid, arginine, histidine) at physiologic pH. The pH-specific protonation is critical for the accurate modeling of ligand interactions, as the electrostatic properties change depending on the pH and thus lead to different binding properties. This is exemplified by the different protonation states of histidine within the protein (Kim et al., 2013). After protonation, the protein needs to undergo energy minimization to relax the structure given the new

Table 1. Selected Open Source and Commercial Software for Major Molecular Modeling Tasks Related to Docking

Molecular Modeling Task	Software/Web Server	Open Source*/Commercial	Website
1. Protein Preparation			
Crystal Structure PDB	RCSB	Open Source	https://www.rcsb.org/
	PDBe	Open Source	http://www.ebi.ac.uk/pdbe/
Protein Visualization	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
	Avogadro	Open Source	https://avogadro.cc/
	PyMOL	Commercial	https://pymol.org/2/
	Schrodinger Maestro	Commercial	https://www.schrodinger.com/maestro https://www.ks.uiuc.edu/Research/vmd/
Homology Modeling	VMD	Open Source	https://www.ks.uiuc.edu/Research/vmd/
	NCBI BLAST	Open Source	https://blast.ncbi.nlm.nih.gov/Blast.cgi
	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
	Modeller	Open Source	https://salilab.org/modeller/
	SWISS-MODEL	Open Source	https://swissmodel.expasy.org/
	I-Tasser	Open Source	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
	Rosetta	Open Source	https://www.rosettacommons.org/software
	Phyre2	Open Source	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
Assigning Protein Protonation States	Schrodinger Prime	Commercial	https://www.schrodinger.com/prime
	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
	Schrodinger	Commercial	https://www.schrodinger.com/
Protein Energy Minimization	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
	Schrodinger Protein Preparation Wizard	Commercial	https://www.schrodinger.com/protein-preparation-wizard
	NAMD	Open Source	http://www.ks.uiuc.edu/Research/namd/
2. Ligand Preparation			
Drawing Chemical Structures	Avogadro	Open Source	https://avogadro.cc/
	Chemdraw	Commercial	https://www.perkinelmer.com/category/chemdraw
	Marvinsketch	Commercial	https://chemaxon.com/products/marvin
	OpenEye Omega	Commercial	https://www.eyesopen.com/omega
	Schrodinger LigPrep	Commercial	https://www.schrodinger.com/ligprep
Downloading Established Ligand Databases	DrugBank	Open Source	https://www.drugbank.ca/
	eMolecules	Commercial	https://reaxys.emolecules.com/index.php
	PubChem	Open Source	https://pubchem.ncbi.nlm.nih.gov/
	ZINC	Open Source	https://zinc15.docking.org/
3. Docking			
Binding Site Prediction	COACH	Open Source	https://zhanglab.ccmb.med.umich.edu/COACH/
	GHECOM	Open Source	http://strcomp.protein.osaka-u.ac.jp/ghecom/
	Schrodinger Sitemap	Commercial	https://www.schrodinger.com/sitemap
Docking Protocol	Surfnet	Open Source	https://www.ebi.ac.uk/thornton-srv/software/SURFNET/
	Autodock	Open Source	http://autodock.scripps.edu/
	DOCK	Open Source	http://dock.compbio.ucsf.edu/
	GOLD	Commercial	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
	RosettaLigand	Open Source	https://rosie.graylab.jhu.edu/ligand_docking
	Schrodinger Glide	Commercial	https://www.schrodinger.com/glide
	FlexX	Commercial	https://www.biosolveit.de/FlexX/
	Surflex	Open Source	http://www.jainlab.org/downloads.html
	FRED/HYBRID	Commercial	https://www.eyesopen.com/oedocking

*w.r.t. academic non-commercial endeavors.

electrostatic properties and that crystallization causes artificial packing of the protein to generate crystals.

If a protein target has not yet been crystallized (e.g., not found in the PDB), then either homology modeling or *ab initio* modeling is required. Homology modeling is the process of predicting a protein's structure from a related homologous protein (also called the template) whose structure has already been determined. Generally, this requires aligning the amino

acid sequence of the target against a large database of known proteins with established structures, usually by a program such as NCBI BLAST (Johnson et al., 2008). A rank-order list of templates based on percent sequence identity is then retrieved, and those with > 60% identity can be considered good templates. Sophisticated homology modeling software tools such as SWISS-MODEL (Waterhouse et al., 2018) or Modeller (Sali and Blundell, 1993; Webb and Sali, 2016) use

RESEARCH TECHNIQUES MADE SIMPLE

additional target sequences of related proteins and multiple templates to predict the best alignment (profile-profile based alignment) and predict the target structure of the protein (Table 1). In contrast, *ab initio* modeling is the process of predicting the tertiary folded three-dimensional structure of a protein from just its amino acid sequence. This is known as the classical “protein folding problem” (Dill et al., 2008) and is an incredibly difficult task. A detailed discussion of homology and *ab initio* modeling is outside the scope of this article; the reader is encouraged to peruse the following articles (Bonneau and Baker, 2001; Hardin et al., 2002).

After target preparation, the ligand(s) must be similarly prepared. Ligands may be individually drawn (two-dimensional [2D] or three-dimensional [3D] structure) using chemical drawing software such as Avogadro (Hanwell et al., 2012). Most commonly, ligands would be downloaded from a vendor or propriety or publicly available databases such as ZINC (Sterling and Irwin, 2015) and PubChem (Kim et al., 2019). Ligand structures (atom types, atom charges, atom connectivities, and spatial coordinates) are represented in a multitude of file types (e.g., SMILES, SDF, MAE, MOL2). While each file type is recognized by different molecular modeling software, they all essentially represent the same information for a given ligand. Once the ligands are drawn or downloaded, 2D structure representations must be converted to 3D. Undefined chiral centers and geometric configurations must be enumerated. The proper protonation states must also be assigned at the appropriate pH, just like the protein target. In addition, ligand tautomers need to be generated. Tautomers are constitutional isomers of a ligand, which occur because of the relocation of a proton. As tautomers can interconvert, it is important to generate feasible tautomers to achieve reliable results in molecular modeling. Depending on the docking algorithm, it is sufficient to have one energy-minimized 3D structure for each ligand representation (protonation state tautomer, stereoisomer), or an ensemble of energetically feasible conformers need to be generated for each representation. Several tools can be used for ligand preparation, for example LigPrep from Schrodinger and Omega from OpenEye (Table 1) to generate high-quality conformer libraries. The reader is referred to Brink and Exner (Brink and Exner, 2009) for a detailed discussion of how docking results are affected by ligand representation.

Docking may then be performed once the protein target and ligands are prepared. The binding site must first be identified. Various software tools can do this using different methods. For example, if a reference ligand (e.g., a known inhibitor) is co-crystallized with the protein target, then the binding site may be considered as the residues located within a set distance from the reference. If no reference is known, a set of residues may be selected to define the pocket. Software, such as Surfnets and SiteMap, aid in binding site identification. A *grid* must then be generated that encompasses the binding site. The center of the grid may be placed at the centroid of the reference ligand or the binding site-defining amino acids. The grid serves as the 3D volume that the docking algorithm uses to place ligands and explore and score binding interactions. The grid includes all definitions of the protein binding site parameters to calculate a docking score to

estimate relative binding affinity. Once the grid is set, docking can then be initiated.

Two general docking algorithms that require slightly different workflows and tools exist: flexible and rigid docking. Here, flexible and rigid refer to the ligand; the protein in most algorithms is kept rigid allowing no or only minimal conformational changes. Flexible docking is used to assess changes in ligand geometry after the binding complex is formed. Thus, the algorithm explores the conformational space of the ligand while docking (i.e., keeps the ligand flexible). It therefore requires only one (energetically favorable) 3D conformation as input per ligand representation. However, given the large degrees of freedom due to conformational sampling, the computational running time is longer than rigid docking. In rigid docking, the 3D ligand is rotated and translated during docking, but its 3D conformation is not changed (i.e., the internal geometry is held rigid). This algorithm therefore requires a pre-generated list of all feasible 3D conformers for each ligand representation; depending on the ligand, typically between 20 and 200 conformers with the goal of obtaining at least one potentially correct conformer. The rigid docking algorithm is much faster but does not consider the flexibility of the ligand within the binding pocket, leading to false-positives and false-negatives. Examples of the best flexible docking tools include Glide (Friesner et al., 2004) from Schrodinger and FlexX (Rarey et al., 1996) from BioSolveIT, and the best rigid docking algorithms include FRED and HYBRID from OpenEye (McGann, 2012). The open source docking software DOCK from UCSF (Kuntz et al., 1982) includes algorithms for rigid and flexible docking. Importantly, both methods ignore protein flexibility in the interest of efficiency, which contributes to decreased accuracy. The consideration of both ligand and protein flexibility upon binding is necessary and is incorporated in methods such as Induced-Fit Docking (Sherman et al., 2004) but at a greater computational cost.

After docking is completed, a file is returned that contains the 3D structure of the docked ligand in the binding site of the target. A table of docking scores to estimate relative binding affinity and, typically, various other energy terms is also provided. The investigator can rank-order the ligands based on the docking score and other considerations and inspect their predicted binding poses and residue interactions. Generally, the more negative the docking score (which is an estimate of relative binding ΔG), the more stable the interaction, and hence the more likely that binding event will occur (Kitchen et al., 2004). From this rank-ordered list, usually the top 10–100 ligands would be obtained for biological testing.

FURTHER CONSIDERATIONS

While docking is widely utilized, it can be a very nuanced methodology. A priori knowledge of the structural biology of the target is necessary for setting up docking experiments. The choice of the initial conformation depends on what the investigator is trying to accomplish (e.g., discovery of agonists, antagonists, or allosteric modulators). In addition, numerous docking algorithms exist that differ in the way they calculate the energetics of the protein-ligand pose and thus rank the compounds and poses. These methods are called

scoring functions and are developed based on different methods, and, thus, no universal scoring function exists. For example, empirical scoring functions are regression-derived equations trained from series of protein-ligand complexes with known binding affinities. The training set used limits the generalized applicability of a particular scoring function. Docking results are thus impacted by false-positives but still help to reduce the vast chemical space for biological testing. A consensus scoring method may be used as a workaround where multiple scoring functions are used for a docking protocol, and molecules that are consistently top-ranked across all scoring functions are then selected for biological testing (Poli et al., 2016). The reader is directed to (Chen, 2015) for a more nuanced discussion regarding the limitations of docking. In general, docking scores are the most useful for the relative ranking of docked compounds and poses in the same or closely related protein and binding site. Docking scores are not the actual binding affinity or binding free energy.

Furthermore, the solvent and salt content (e.g., water and concentration of sodium, chloride, and potassium ions) are important contributors to the binding ΔG . Most docking programs perform simulations within a vacuum without accounting for physiologic solvent and salt content. However, the solvent effect is implicitly modeled in the scoring function. Thus, more accurate solvent models are typically not used in primary scoring, because of the high computational cost (hardware and time requirements) of performing these simulations accurately. Thus, typically only a few (< 1000) molecules would be chosen after docking for subsequent refinement under solute and solvent considerations. These post-docking energy calculations help to increase the true positive rate and are known as molecular mechanics generalized Born molecular surface area and molecular mechanics generalized Poisson–Boltzmann surface area methods, which are reviewed in (Kerrigan, 2013).

It is also important to note, however, that docking only represents a single “snapshot” of the molecular interaction. Biological systems are always in flux, and their interactions change over time. Recent technological advances now allow for the highly granular study of molecular interactions at the atomic level including solvent. Specifically, molecular dynamics is the study of atomic interactions over time as they depend on Newtonian mechanics, temperature, and physical forces. The reader is referred to (Hospital et al., 2015) on the application of molecular dynamics for drug discovery.

CONCLUDING REMARKS

Drug discovery is a pain-staking process requiring enormous research and development investment; yet success rates are low. With increasing amounts of available data (big data) and huge advances in computational hardware and algorithms, computational tools are now used throughout the drug discovery pipeline and have the potential to aid in revitalizing the pharmaceutical pipeline. One such computational method in the preclinical hit and lead identification stage is molecular docking, which is now scalable to hundreds of millions of compounds. As actionable protein targets are becoming better elucidated in dermatologic conditions,

MULTIPLE CHOICE QUESTIONS

- Why is docking useful for early stage preclinical drug discovery?
 - Enriching very large chemical databases so as to test compounds with greatest likelihood of interacting with your target
 - Understanding the specific protein-ligand atomic interactions (i.e., binding mode) to guide lead optimization for example
 - A & B
 - None of the above
- What is a major limitation of docking?
 - It can screen only a small set of compounds against a protein target
 - No one docking software/algorithm can be used for all protein target systems
 - It is unable to study how a drug interacts with its target over time
 - B & C
- Which of the following is NOT required for docking?
 - Protein structure
 - Ligand Structure
 - Reference structure of protein-ligand interaction
 - None of the above
- Docking may be applied for virtual screenings of molecule libraries of what size?
 - Hundreds
 - Thousands
 - Millions
 - Any size library
- You have already discovered a potential new drug for your target. Now you are interested in predicting the potential off-target effects of your drug. What is required for the application the docking for your drug in predicting off-target effects?
 - Three-dimensional structure of your drug
 - Library of potential human protein target structures
 - Library of cytochrome P450 target structures
 - All of the above

Note: See online version of this article for a detailed explanation of correct answers.

molecular modeling, such as docking, has great potential to aid in expediting the process from drug ideation to clinical use.

ORCID

Naïem T. Issa: <https://orcid.org/0000-0002-5885-3909>
 Evangelos V. Badiavas: <https://orcid.org/0000-0002-3664-3409>
 Stephan Schürer: <https://orcid.org/0000-0001-7180-0978>

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by NIH grants U54HL127624 (BD2K LINCS Data Coordination and Integration Center, DCIC) and U24TR002278 (Illuminating the Druggable Genome Resource Dissemination and Outreach Center, IDG-RDOC). The BD2K LINCS DCIC is awarded by the National Heart, Lung, and Blood Institute through funds provided by the trans-NIH Library of Integrated Network-based Cellular Signatures (LINCS) Program (<http://www.lincsproject.org/>) and the trans-NIH Big Data to Knowledge (BD2K) initiative (<https://datascience.nih.gov/bd2k>). The IDG-RDOC (<http://druggablegenome/>) is a component of the Illuminating the Druggable Genome (IDG) project (<https://commonfund.nih.gov/idg>) awarded by the National Center for Advancing Translational Sciences (NCATS). LINCS, IDG, and BD2K are NIH Common Fund projects. We sincerely thank Open Eye (OE) for a free academic license to their software tools, including the OE docking.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to this paper. Teaching slides are available as supplementary material.

REFERENCES

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- Bonneau R, Baker D. Ab initio protein structure prediction: progress and prospects. *Annu Rev Biophys Biomol Struct* 2001;30:173–89.
- Chen YC. Beware of docking! *Trends Pharmacol Sci* 2015;36:78–95.
- Chenera B, Desjarlais RL, Finkelstein JA, Eggleston DS, Meek TD, Tomaszek TA, et al. Nonpeptide HIV protease inhibitors designed to replace a bound water. *Bioorg Med Chem Lett* 1993;3:2717–22.
- Desjarlais RL, Dixon JS. A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. *J Comput Aided Mol Des* 1994;8:231–42.
- Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys* 2008;37:289–316.
- Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 2016;11:905–19.
- Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47:1739–49.
- Ghosh S, Sinha M, Bhattacharyya A, Sadhasivam S, Megha J, Reddy S, et al. A rationally designed multifunctional antibiotic for the treatment of drug-resistant acne. *J Invest Dermatol* 2018;138:1400–8.
- Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* 2012;4:17.
- Hardin C, Pogorelov TV, Luthey-Schulten Z. Ab initio protein structure prediction. *Curr Opin Struct Biol* 2002;12:176–81.
- Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular Dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* 2015;8:37–47.
- Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Molec Graph* 1996;14(33–8):27.
- Hussein HA, Geneix C, Petitjean M, Borrel A, Flatters D, Camproux A. Global vision of druggability issues: applications and perspectives. *Drug Discov Today* 2017;2:404–15.
- Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;36(Web Server issue):W5–9.
- Kerrigan JE. Molecular Dynamics simulations in drug design. *Methods Mol Biol* 2013;993:95–113.
- Kim MO, Nichols SE, Wang Y, McCammon JA. Effects of histidine protonation and rotameric states on virtual screening of *M. tuberculosis* RmlC. *J Comput Aided Mol Des* 2013;27:235–46.
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;47:D1102–9.
- Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;3:935–49.
- Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982;161:269–88.
- Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS, et al. Ultra-large library docking for discovering new chemotypes. *Nature* 2019;566:224–9.
- Mann T, Gerwat W, Batzer J, Eggers K, Scherner C, Wenck H, et al. Inhibition of human tyrosinase requires molecular motifs distinctively different from mushroom tyrosinase. *J Invest Dermatol* 2018;138:1601–8.
- McGann M. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des* 2012;26:897–906.
- Mirza SB, Lee RCH, Chu JH, Salmas RE, Mavromoustakos T, Durdagi S. Discovery of selective dengue virus inhibitors using combination of molecular fingerprint-based virtual screening protocols, structure-based pharmacophore model development, molecular dynamics simulations and in vitro studies. *J Mol Graph Modell* 2018;79:88–102.
- Mirza SB, Salmas RE, Fatmi MQ, Durdagi S. Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study. *J Mol Graph Model* 2016;66:99–107.
- Poli G, Martinelli A, Tuccinardi T. Reliability analysis and optimization of the consensus docking approach for the development of virtual screening studies. *J Enzyme Inhib Med Chem* 2016;31(Suppl.2):167–73.
- Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996;261:470–89.
- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
- Sevrioukova IF, Poulos TL. Current approaches for investigating and predicting cytochrome P450 3A4-ligand interactions. *Adv Exp Med Biol* 2015;851:83–105.
- Sherman W, Beard HS, Farid R. Use of an induced fit receptor structure in virtual screening. *J Med Chem* 2004;67:83–64.
- Sterling T, Irwin JJ. Zinc 15 – ligand discovery for everyone. *J Chem Inf Model* 2015;55:2324–37.
- ten Brink T, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein–ligand docking results. *J Chem Inf Model* 2009;49:1535–46.
- Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multi-threading. *J Comput Chem* 2010;31:455–61.
- van Eijl S, Zhu Z, Cupitt J, Gierula M, Götz C, Fritsche E, et al. Elucidation of xenobiotic metabolism pathways in human skin and human skin models by proteomic profiling. *PLOS ONE* 2012;7:e41721.
- Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys* 2016;18:12964–75.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296–303.
- Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 2016;13:6.1–5.6.30.
- Zhao C, Yang SH, Khadka DB, Jin Y, Lee KT, Cho WJ. Computer-aided discovery of aminopyridines as novel JAK2 inhibitors. *Bioorg Med Chem* 2015;23:985–95.

DETAILED ANSWERS

1. Why is docking useful for early stage preclinical drug discovery?

Answer: C. Molecular docking in drug discovery serves two major functions. The first is to provide an efficient chemical screening method (“virtual screening”) of hundreds of thousands to tens or even hundreds of millions of compounds and narrow them down to a small number of compounds with the greatest likelihood of interacting with the protein target in question. This enriched set of compounds can then be purchased from appropriate vendors for biological testing. Thus, docking optimizes the preclinical drug discovery pipeline to enrich for compounds that will have the greatest likelihood of biological success. Second, molecular docking can assist in determining the molecular interactions of a drug with the target protein at the atomic level. For example, by docking a drug into the binding site of its known protein target, the experimenter can study the interactions the drug makes with the amino acid side chains of the protein. This information can be further leveraged to find novel drugs with improved binding affinity and possibly better side effect profiles by avoiding binding to undesired off-targets (via docking into these proteins). This is typically done through medicinal chemistry efforts where chemical moieties may be added or changed to enforce or prevent certain interactions. For example, the addition of a carboxyl group (-COOH) to a drug may strengthen its binding by contributing an additional hydrogen bond to an amino acid sidechain, since it becomes deprotonated into a carboxylate ion (-COO⁻) if its pKa is higher than that of physiologic pH, thus increasing the binding affinity. Alternatively, the addition of bulky groups, such as a cyclohexyl group, may cause steric hindrance and decrease the ability of the drug to fit in the binding pocket, thus decreasing the binding affinity.

2. What is a major limitation of docking?

Answer: D. Various docking software and/or algorithms exist that employ diverse scoring functions for calculating the free energy of binding (e.g., forcefield, empirical, knowledge-based, and target-based scoring functions). As such, there is no one universal method for employing docking to a given protein target system. However, one can combine multiple docking methods in a consensus approach, termed “consensus docking”, and enrich for drugs that consistently are ranked highest across the majority of docking methods. In addition, docking provides a single snapshot in time of a drug interacting with the binding pocket of the protein target. Docking is unable to elucidate how the drug will continue to interact within the pocket over time, or if the drug will induce large-scale conformational changes in the

protein target. These questions are best answered through other methods such as molecular dynamics (MD) simulations.

3. Which of the following is NOT required for docking?

Answer: C. Molecular docking requires a model structure of the protein target of interest (e.g., crystal structure or homology model) and 3D structures of the ligands to be docked. The protein structure needs to be energy-minimized, and the protein amino acid side chains should have the appropriate protonation states based on the pH of the biological system (e.g., the software should be told to prepare the protein structure at a physiological pH of 7.4). Similarly, ligands should be prepared at the appropriate pH, as the protonation state changes their physicochemical features and thus binding properties. While it is desirable to have a reference protein structure with a bound ligand (e.g., inhibitor) as an example binding site/mode, it is not necessary. The experimenter may use computer software to identify one or more putative binding sites to serve for subsequent docking studies.

4. Docking may be applied for virtual screenings of molecule libraries of what size?

Answer: D. Docking is versatile in its use for studying the potential mechanism of action for a single drug molecule and for screening molecule libraries that can be as large as hundreds of millions of compounds. The major limitation in screening large libraries is the computational cost. Many academic centers or industry entities have dedicated computational infrastructures to allow for high-throughput efficient docking of very large molecule libraries.

5. You have already discovered a potential new drug for your target. Now you are interested in predicting the potential off-target effects of your drug. What is required for the application the docking for your drug in predicting off-target effects?

Answer: D. Docking is frequently used to predict off-target effects of a drug, as well as predicting its pharmacokinetics with respect to drug metabolism. Investigators often perform an *inverse docking* procedure where a drug of interest is docked against a large library of human protein target structures. Those interactions that score well can lead the investigator to potentially infer the off-target effects and its consequence on human biology. In addition, docking against cytochrome P450 enzyme structures allows the investigator to predict important interactions that will affect how the drug is metabolized into other potentially active metabolites. The investigator may also model these metabolites and predict their targets using docking.