



Review

Pharmacological Chaperones and Protein Conformational Diseases: Approaches of Computational Structural Biology

Daniela Grasso , Silvia Galderisi, Annalisa Santucci and Andrea Bernini *

Department of Biotechnology, Chemistry, and Pharmacy, University of Siena, 53100 Siena, Italy

* Correspondence: andrea.bernini@unisi.it

Abstract: Whenever a protein fails to fold into its native structure, a profound detrimental effect is likely to occur, and a disease is often developed. Protein conformational disorders arise when proteins adopt abnormal conformations due to a pathological gene variant that turns into gain/loss of function or improper localization/degradation. Pharmacological chaperones are small molecules restoring the correct folding of a protein suitable for treating conformational diseases. Small molecules like these bind poorly folded proteins similarly to physiological chaperones, bridging non-covalent interactions (hydrogen bonds, electrostatic interactions, and van der Waals contacts) loosened or lost due to mutations. Pharmacological chaperone development involves, among other things, structural biology investigation of the target protein and its misfolding and refolding. Such research can take advantage of computational methods at many stages. Here, we present an up-to-date review of the computational structural biology tools and approaches regarding protein stability evaluation, binding pocket discovery and druggability, drug repurposing, and virtual ligand screening. The tools are presented as organized in an ideal workflow oriented at pharmacological chaperones' rational design, also with the treatment of rare diseases in mind.

Keywords: pharmacological chaperones; protein conformational diseases; computational structural biology; protein stability; transient pockets; pocket druggability; drug repurposing; virtual screening; molecular docking



Citation: Grasso, D.; Galderisi, S.; Santucci, A.; Bernini, A.

Pharmacological Chaperones and Protein Conformational Diseases: Approaches of Computational Structural Biology. *Int. J. Mol. Sci.* **2023**, *24*, 5819. <https://doi.org/10.3390/ijms24065819>

Academic Editor: Andrzej Kloczkowski

Received: 5 February 2023

Revised: 9 March 2023

Accepted: 16 March 2023

Published: 18 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Protein Conformational Diseases

Several biological mechanisms for the function of our organism rely on proteins, and a protein's biological function is determined by its three-dimensional structure. Therefore, these molecules must be folded into a native state to function correctly. Still, the process of achieving native conformation is complex, which makes identifying the mechanism responsible for protein assembly central to structural biochemistry investigation. Indeed, an error in such a process could result in a misfolded protein or a missed quaternary assembly, leading to pathological effects. Known to cause cellular malfunctioning or death, altered proteins lose their function and lead to deleterious forms or aggregates [1,2]. Many factors contribute to the loss of the proper three-dimensional structure. Still, the amino acid composition remains the primary factor since mutations may destabilize the correct fold of a protein or stabilize a misfolded state [3–5].

Among protein-altering mutations, missense mutations are the most common [6]. A missense mutation occurs when one nucleotide base is substituted for another in a DNA sequence, resulting in a different codon and amino acid. The prevalence of missense mutations in any given gene is estimated at 2% [7]. Recent studies have focused on protein misfolding caused by missense mutations, which has led to the association with several serious diseases [4].

When proteins cannot fold into a native structure, a profound detrimental effect is expected, often resulting in disease. Known as protein conformational diseases, such disorders arise when proteins adopt abnormal conformations. Many people suffer from

conformational diseases, significantly challenging human health [8]. The present review focuses on examples of rare types of conformational diseases, such as (with involved protein in parenthesis): Phenylketonuria (phenylalanine hydroxylase, PAH) [9,10], Alkaptonuria (1,2 homogentisate dioxygenase, HGD) [11,12], Transthyretin-related hereditary amyloidosis (transthyretin, TTR) [13], and the group of lysosomal storage diseases [14] including GM1-gangliosidosis (β -galactosidase), Tay-Sachs disease (β -hexosaminidase A), Sandhoff disease (β -hexosaminidases A and B), AB variant of GM2-gangliosidosis (GM2 activator protein), Fabry disease (α -galactosidase A), Gaucher disease (β -glucocerebrosidase), Pompe disease (α -glucosidase), mucopolysaccharidosis IIIC (heparan- α -glucosaminide N-acetyltransferase), and Batten disease (battenin).

Molecular mechanisms that underlie many protein conformational diseases are very complex and multifactorial. Several phenomena are associated with conformational diseases, including loss of function, toxic gain of function, improper degradation, and improper localization. A loss-of-function mutation results in an impairment of the biological function, either complete (amorphic), which is analogous to protein null mutations, or partial (hypomorphic). Often, loss-of-function results from premature stop codons due to nonsense or frameshift mutations. The deleterious mRNA obtained usually produces no protein rather than a truncated polypeptide [15]. Hence, such mutations are usually of poor interest to the structural biologist, who is keener on missense mutations that change the amino acid sequence and result in protein products, although misfolded or non-functional. In structural terms, such mutations lead to (local) misfolding, active site disruption, or quaternary structure disruption. On the opposite side, the strengthening of protein–protein interactions induced by amino acidic mutation is a gain-of-function often resulting in a toxic effect. Mutations may increase protein activity (hypermorphic) or introduce a new function (neomorphic). Improper degradation relates to cellular degradation systems, such as autophagy; despite being essential for preventing the accumulation of misfolded proteins, it can sometimes lead to disease by over-actively degrading proteins that, despite their mutations, retain some functionality. The mislocalization of proteins arises from mutation destabilizing the correct fold and preventing the trafficking to the native subcellular compartment [16].

From a structural biology point of view, conformational diseases arise from a protein impairment to reach its native three-dimensional structure or stability. In turn, restoring the correct protein structure would rescue the biological activity and mitigate the disease; if molecules can mediate such a process, therapy development can be envisaged, but can protein rescue be mediated by molecular species? It happens in physiological conditions, where the acquisition of native conformation is made possible by specialized macromolecules called chaperones, which help proteins find their native functional conformation.

The chaperones to which we usually refer are proteins that assist the folding of other macromolecular structures and prevent the unwanted associations of unfolded polypeptide chains, both in physiological conditions and stress conditions. Since the tendency of a protein to aggregate increases in stress conditions, some chaperones are heat shock proteins (HSP), expressed in response to high temperatures or cellular stress. Physiological chaperons are essential in most cellular compartments, especially where proteins are subjected to potential stress, for example, mitochondrion, because of the low pH and the presence of degradation enzymes.

Based on such mechanisms, a new class of drugs is being exploited for protein rescue in conformational diseases: the pharmacological chaperones (PC).

2. Pharmacological Chaperones

Pharmacological chaperones (PC) are small molecule drugs designed to facilitate the correct folding of a protein and to re-establish its functionality [17,18] (see Figure 1).

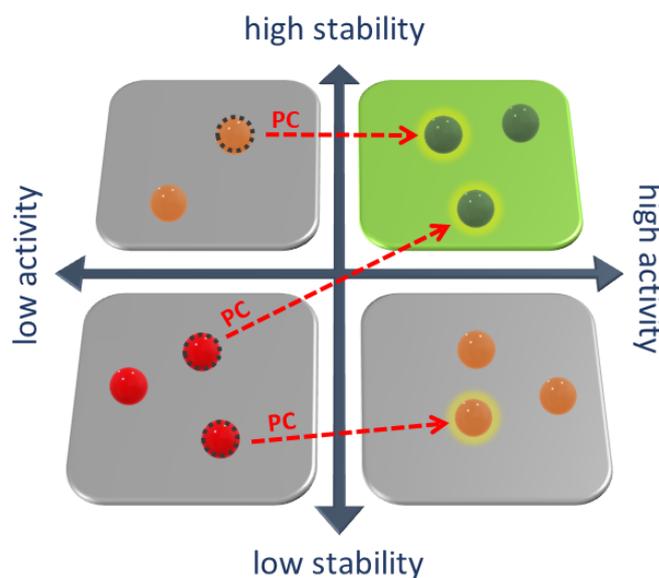


Figure 1. Native state protein has high activity and stability (green sphere). Misfolded proteins have lower stability, activity (orange spheres), or both (red). Pharmacological chaperones, binding with misfolded protein, increase its stability and activity. After binding with chaperones, proteins with high residual activity (**upper left square**) assume stability and activity similar to the wild-type. Proteins with lower residual activity (**lower left square**) assume high stability after binding with pharmacological chaperones, increasing the availability of protein in the cell.

Similarly to physiological chaperones, these small molecules bind to poorly folded proteins, bridging non-covalent interactions (hydrogen bonds, electrostatic, and van der Waals) that were lost or loosened due to, e.g., missense mutations, thus strengthening the protein structure [19]. In addition, the low molecular weight of the PC gives the same advantages as other classes of drugs, such as molecular trafficking through membranes, reaching a variety of tissues [20], and oral availability [21].

Pharmacological chaperones should not be confused with chemical chaperones, such as trehalose and glycerol. The chemical chaperones are non-specific molecules; therefore, they bind to and stabilize all proteins. The unspecific binding and the required high concentration levels cause them to be unsuitable for therapeutic use.

Based on their molecular interaction with their target protein, PCs are classified into different types, namely, competitive inhibitors, enzyme cofactors, allosteric ligands, and alternative binders, described as follows.

2.1. Competitive Inhibitors

Many of the PCs described so far have been competitive inhibitors of lysosomal storage disorders (LSDs) [22]. They create noncovalent hydrogen-bonding networks and van der Waals interactions in the same pocket as the enzyme substrate, stabilizing the protein structure. Consequently, when treating patients, a balance must be found between folding enhancement and enzyme inhibition by using subinhibitory concentrations. Moreover, PCs show a pH-dependent affinity to their target protein. The PC binds reversibly to its target protein in the endoplasmic reticulum (ER) at neutral pH, stabilizing and facilitating its transportation. An acidic pH in the lysosome allows dissociation between PC and its target protein, enabling metabolization. Dissociation at acidic pH is, therefore, crucial to broaden the narrow therapeutic index between enzymatic recovery and inhibition. Such behavior has been exploited to develop a PC for Fabry disease. Fabry disease is caused by variants in the GLA gene, coding for α -galactosidase A. Such enzyme, active in lysosomes, breaks down the fatty substance globotriaosylceramide, a glycosphingolipid. As a result of loss-of-function variants in the GLA gene, globotriaosylceramide builds up in cells throughout the body, particularly cells lining blood vessels in the skin and cells in the

kidneys, heart, and nervous system. Fabry disease is characterized by the progressive accumulation of this substance in cells that results in various signs and symptoms. A low molecular weight iminosugar analog of the terminal galactose residue of GL-3, called migalastat, binds selectively and reversibly to the active site of various variants of α -galactosidase A enzyme [23–25]. Through this binding mechanism, and at levels below the inhibitory concentration, migalastat acts as a pharmacological chaperone, stabilizing variants of α -galactosidase A in the ER and facilitating proper trafficking to lysosomes [26]. Once in lysosomes, migalastat dissociates from α -galactosidase A due to the lower pH, allowing the enzyme to break down GL-3 [24–26]. Upon dissociation from the enzyme, migalastat is rapidly excreted from the body. Another example of a competitive inhibitor is iminosugar isofagomine, targeting and stabilizing mutants of β -glucosidase [27] and acting as PC in Gaucher disease [28].

2.2. Enzyme Cofactors

Enzyme cofactors are another type of PC. It may be beneficial for stabilizing misfolded proteins to increase the amount of the natural cofactor of an enzyme. The most common example of IEM is tetrahydrobiopterin (BH4), the cofactor for phenylalanine hydroxylase, which is defective in patients with phenylketonuria (PKU) [29–31]. About half of PKU patients respond well to this treatment [18], and BH4 has received FDA approval and is marketed under the commercial name Kuvan (see Table 1).

2.3. Allosteric Ligands

PCs can also take the form of allosteric ligands. These PCs stabilize their targeted proteins by interacting in pockets other than the active site without impairing the function of the proteins, a clear advantage over competitive inhibitor PCs. A renowned example is the allosteric ligand Tafamidis (see [32–35] and Figure 2), already approved for transthyretin-related hereditary amyloidosis (ATTR) [36] treatment.

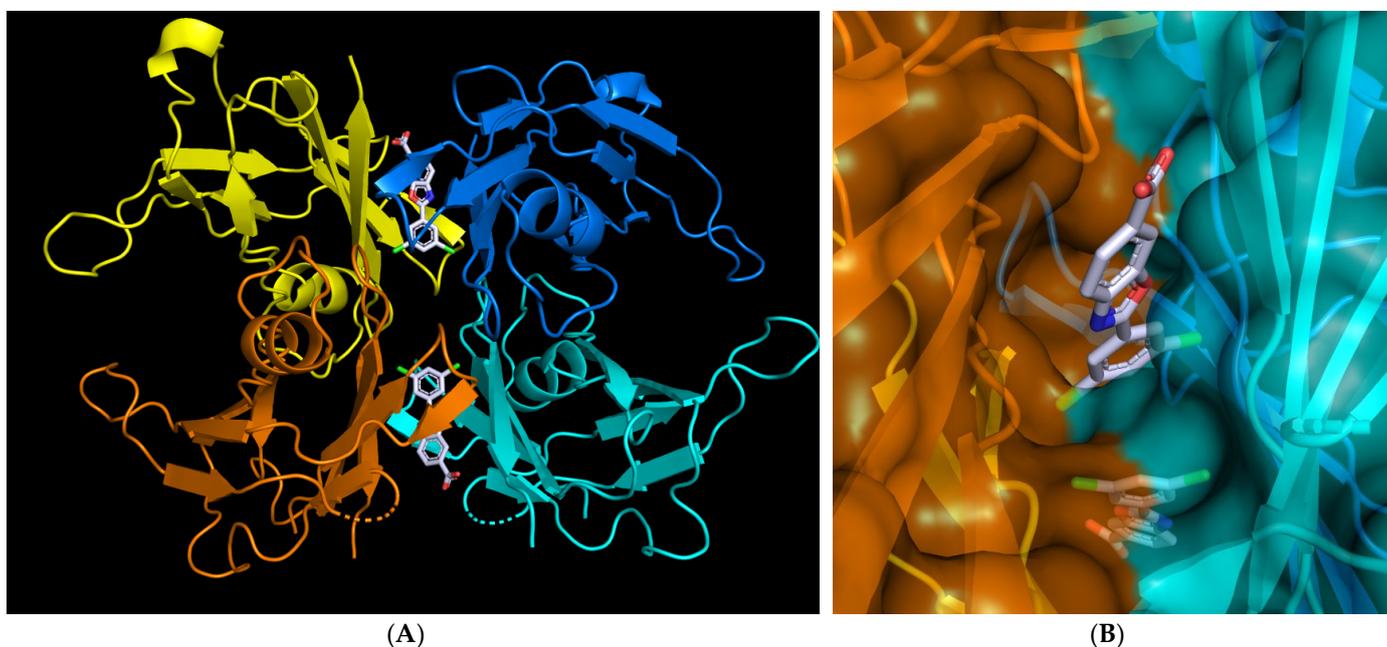


Figure 2. Three-dimensional structure of tafamidis-bound transthyretin (TTR). (A) The TTR homotrimer, arranged as a dimer of dimers (colored in yellow/orange and cyan/blue), is represented as ribbons; subunits assemble around a central channel that accommodates two drug molecules at the weak interacting surface of the dimers. (B) Close view of a tafamidis molecule filling the gap between surfaces of two dimers. Representations are made with PyMol using a structure from Protein Data Bank (PDB ID 3TCT, accessed on 10 January 2023).

ATTR is an autosomal neurodegenerative disease characterized by neuropathy and cardiomyopathy caused by variants in the transthyretin (TTR) gene. TTR carries vitamin A (retinol) and the hormone thyroxine (T4) throughout the body. It has a homotetrameric structure arranged as a dimer of dimers. In the presence of conformational variants, however, the equilibrium is tipped toward misfolded monomers, which are the origins of pathogenic aggregates and cell death. Patients carrying the most common variants of TTR, p.Val30Met, accounting for 85% of cases worldwide, were demonstrated to be protected by Tafamidis [37]. A successive study demonstrated that Tafamidis is also effective on eight variants [38] other than p.Val30Met: p.Asp38Ala, p.Gly47Ala, p.Leu58His, p.Thr60Ala, p.Phe64Leu, p.Ser77Phe, p.Ser77Tyr, and p.Ile107Val. The PC selectively binds to the two largely unoccupied thyroxine-binding sites of the tetramer located at the weak dimer–dimer interface (Figure 2), incrementing the number of interactions between protomers. Thus, in tafamidis-bound TTR, the weaker dimer–dimer interface is stabilized against dissociation, which is the rate-limiting step in forming amyloids [35]. Consequently, Tafamidis slows the progression of the disease. Many other conformational variants of the TTR protein have been treated with Tafamidis [39–42]. Tafamidis received approval from FDA and is marketed as Vyndamax (see Table 1).

Table 1. Pharmacological chaperones as therapeutics.

Name of Disease	Pharmacological Chaperones	Clinical Status
Transthyretin-related hereditary amyloidosis	Vyndamax (Tafamidis)	Market approved
Phenylketonuria	Kuvan (tetrahydrobiopterin or BH4)	Market approved
Fabry disease	Migalstat (1-deoxygalactonojirimycin or Galafold)	Market approved
Gaucher disease, Type 1	Afegostat tartrate (Isogomine or AT2101)	Phase 2 NCT00446550
Gaucher disease, Type 1	Ambroxol	Phase 2 NCT03950050
Gaucher disease, Type 1	NCGC607	Preclinical cell-based study [43]
Pompe disease	Duvoglustat	Phase 2 NCT00688597
Pompe disease (late-onset)	Miglustat (AT2221) (with alglucosidase alfa, ATB200)	Phase 3 NCT03729362
Gangliosidoses, GM1	N-octyl 4-epi-β-valienamine	Preclinical in vivo study [44]
Gangliosidoses, GM1	1,5-dideoxy-1,5-iminoribitol C-glycoside	Preclinical cell-based study [45]
Gangliosidoses, GM2		
Sandhoff disease	Pyrimethamine	Phase $\frac{1}{2}$ NCT01102686
Tay-Sachs disease		
Mucopolysaccharidosis IIIC	Glucosamine	Preclinical in vivo study [46]
Batten disease	CS38	Preclinical cell-based study [47]

2.4. Alternative Binders

Research is becoming increasingly focused on finding binding sites other than the active ones to avoid interfering with the same biological function subject of the rescue strategy. In recent years, computational methods have been used for discovering and examining new hotspots that are not necessarily associated with active spots. Modulating the protein–protein interface by pharmacological chaperones is also becoming increasingly important in drug discovery, especially for diseases related to quaternary protein instability [48]. Such points will be addressed in the “Pocket prediction tools” and “Exploiting transient pockets” sections.

Several other diseases are treated with pharmacological chaperones, with molecules already marketed and others under clinical trials of various phases. Conformational diseases and the PC targeted against are reported in Table 1. As for drug development in medicinal chemistry, a broad-spectrum screening of chemical libraries (high throughput screening, HTS) [49] is the faster and more efficient way to identify new-generation chaperones. However, PC development is often targeted at rare and neglected diseases, for

which the development of new drugs is too long and expensive, with a risk of failure. For this reason, in the last years, the evolution of the drug repositioning approach has been proposed as a more suitable and less expensive way to identify lead compounds in rare and neglected disease drug development.⁴

3. Drug Repositioning

Drug repositioning (DR), also known as drug repurposing or retasking, is finding new indications for known drugs outside the scope of the original use [50]. The origin of drug repositioning derives from the capacity of some compounds to perturb multiple biological targets, allowing some drugs to be exploited for more than one disease [51,52].

DR usually focuses on molecules that have cleared phase I safety trials but have yet to show efficacy for the intended indication. Therefore, DR can probably skip the preclinical and phase I study [53], which can reduce the cost throughout drug development. The cost of repurposed drugs is typically reduced (50–60%), and their approval generally is quicker (3–5 years) [54,55]. Moreover, repurposed drugs achieve market approval at a higher rate than new drugs, giving companies an incentive to repurpose existing assets [54,56]. Nowadays, repositioning previously approved drugs is a common practice because of the increasing number of computational methods that reduce the number of potential compounds to be screened with high precision and sensitivity [51]. Therefore, this new method offers more significant advantages than *de novo* drug discovery, which is costly and time-consuming, sometimes with a scarce chance of success.

Drug repositioning finds a convenient use *in silico*. Computational methods provide a valuable way for potential drug candidates to be identified and their interactions with proteins to be predicted. Such a method is called molecular structured-based virtual screening. It is commonly used in drug discovery by testing a large number of compounds and calculating the affinity with a target site. Further, it is known that many molecules have an affinity with more than one protein; thus, by screening known drugs, molecules active on targets of other diseases can possibly be found [51]. From a DR perspective, such virtual screening can be exploited for the out-licensing of drugs active on targets other than the original. The union of virtual screening and drug repositioning has been a significant step forward in drug discovery, especially for rare diseases [57].

The molecules that could be used for drug repositioning are [52]: (i) drugs in clinical trials that have a common mechanism of action with another disease; therefore, clinical trials for the new indication could be conducted in parallel with those for the original indication; (ii) drugs failed in a particular indication in clinical development without safety problems; (iii) drugs taken off the market for commercial reasons; (iv) drugs with patent near to expire; and (v) discovered drugs not launched in markets of developed areas.

Ambroxol, a well-known molecule marketed as expectorant, is a case study of a drug repositioned as PC. In 2009 Maegawa et al. [58] carried out a library screening of FDA-approved drugs in search of stabilizers of variants of glucocerebrosidase (GCase), the deficient enzyme at the base of Gaucher disease (GD), the most prevalent lysosomal storage disease. Ambroxol was found to be a pH-dependent, mixed-type inhibitor of GCase and a stabilizer of the p.Asn370Ser and p.Phe213Ile variants. A few years later, the drug was proven effective against Gaucher variants [59]. Ambroxol improved or arrested the progression of neurological symptoms, such as dystonia and gait disturbances, in a clinical trial of Gaucher disease type 3 patients [60]. Ambroxol has been tested in Phase 2 clinical trial (see Table 1), and a recent observational study [61] has encouraged its off-label usage.

In summary, drug repositioning aims to extend a drug's patent life and administer it to more than one pathology, increasing revenue. Indeed, drug repositioning contributes to improving pharma companies' attention toward rare or neglected diseases. The FDA's Office of Orphan Products Development aims to create strategies for treating and diagnosing rare diseases; also, it maintains the Rare Disease Repurposing Database [62].

4. Predicting the Pathogenicity of Variants and the Effect on Protein Stability

The first step in a bioinformatic workflow for developing a PC for protein rescue in a conformational disease framework is assessing the misfolding resulting from pathological variants and their extent. Many tools are available to accomplish this task, usually web tools (see Table 2) suitable for beginners. They should not be used as black-box anyway, and the results should be checked against experimental evidence and from more advanced simulation tools like molecular dynamics simulation (as illustrated later); also, comparing tool results using different approaches is valuable (see “Method” column in Table 2). For example, SIFT [63] is a multistep algorithm that uses sequence conservation and amino acid properties to predict whether an amino acid substitution is deleterious or neutral. Panther [64] is a valuable software system for predicting genes’ functions based on their evolutionary relationships. SNAP2 [65] is based on a neural network machine-learning algorithm, which distinguishes between neutral and non-neutral SNPs by taking evolutionary information derived from an automatically generated multiple sequence alignment. SNPs&GO [66] is a method that uses a protein sequence to predict whether single-point mutations can lead to diseases in humans. PolyPhen-2 [67] is a trained Naïve Bayes classifier that uses eight sequence-based and three structure-based predictive features automatically selected by its algorithm. FatHMM [68] predicts the functional effects of protein missense mutations by combining sequence conservation within hidden Markov models. VarMod [69] uses structural modeling of proteins, their ligand binding sites, and protein–protein interface sites to analyze non-synonymous single nucleotide variants to identify those that may play a role in diseases. MutPred2 [70] is a machine-learning-based method that integrates genetic and molecular data to predict the pathogenicity of amino acid substitutions and their molecular mechanisms. SNPdryad [71] builds a multiple sequence alignment using only protein orthologs and employs a random forest classifier to derive a deleterious prediction score. ENTPRISE [72] predicts human disease-associated amino acid mutation using a boosted tree regression machine-learning approach from sequence entropy. MutationAssessor [73] predicts the functional impact of amino acid substitutions in proteins, which is assessed based on the evolutionary conservation of the affected amino acid in protein homologs. Mupro [74] uses support vector machine approach combined with a local window centered around the mutated residue as input. CUPSAT [75] uses amino acid atom potentials and torsion angle distribution to assess the amino acid environment of the mutation site and the effect on stability. INPS [76,77] is based on physicochemical descriptors extracted from the protein sequence or structure to evaluate the impact of non-synonymous SNP on protein stability. SuSPect [78] uses 77 sequence-, structure-, and systems biology-based features to train a support vector machine to discriminate between disease-causing and neutral variants. SDM [79] is a computational method that analyzes the variation of amino acid replacements occurring at a specific structural environment for predicting protein stability upon mutation.

Table 2. Protein stability online prediction tools are listed along with the web address and method implemented.

Name	URL	Method
SIFT	https://sift.bii.a-star.edu.sg/ (accessed on 10 January 2023)	Sequence conservation and amino acids properties
Panther	http://pantherdb.org/tools/csnpscoreForm.jsp (accessed on 10 January 2023)	Evolutionary preservation
SNAP2	https://roslab.org/services/snap/ (accessed on 10 January 2023)	Neural networks
SNPs&GO	https://snps-and-go.biocomp.unibo.it/snps-and-go/ (accessed on 10 January 2023)	Support vector machines
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/ (accessed on 10 January 2023)	Naïve Bayes classifier
FatHMM	http://fathmm.biocompute.org.uk/ (accessed on 10 January 2023)	Hidden Markov models
VarMod	http://www.wasslab.org/varmod/ (accessed on 10 January 2023)	Support vector machines
MutPred2	http://mutpred.mutdb.org/ (accessed on 10 January 2023)	Neural networks

Table 2. Cont.

Name	URL	Method
SNPdryad	https://maayanlab.cloud/datasets2tools/landing/tool/SNPdryad (accessed on 10 January 2023)	Sequence alignment using protein orthologs
ENTPRISE	http://cssb2.biology.gatech.edu/ENTPRISE/ (accessed on 10 January 2023)	Sequence entropy and predicted protein structures
MutationAssessor MUpro	http://mutationassessor.org/r3/ (accessed on 10 January 2023) https://mupro.proteomics.ics.uci.edu/ (accessed on 10 January 2023)	Evolutionary preservation Support vector machines
CUPSAT	http://cupsat.tu-bs.de/ (accessed on 10 January 2023)	Amino acid–atom potentials and torsion angle distribution
INPS	https://inpsmd.biocomp.unibo.it/inpsSuite (accessed on 10 January 2023)	Support vector machines
SuSPect	http://www.sbg.bio.ic.ac.uk/~suspect/ (accessed on 10 January 2023)	Support vector machines
SDM	http://marid.bioc.cam.ac.uk/sdm2/prediction (accessed on 10 January 2023)	Graph-based signatures
mCSM-ppi2	https://biosig.lab.uq.edu.au/mcsm_ppi2/ (accessed on 10 January 2023)	Graph-based signatures
DUET	http://biosig.unimelb.edu.au/duet/ (accessed on 10 January 2023)	Support vector machines
mCSM-Membrane	https://biosig.lab.uq.edu.au/mcsm_membrane/ (accessed on 10 January 2023)	Graph-based signatures
mCSM-AB	https://biosig.lab.uq.edu.au/mcsm_ab/ (accessed on 10 January 2023)	Graph-based signatures
DynaMut2	https://biosig.lab.uq.edu.au/dynamut2/ (accessed on 10 January 2023)	Graph-based signatures and normal mode dynamics

Similarly, mCSM-PPI2 [80] uses an optimized graph-based signature approach to evaluate the molecular mechanism of the mutation by modeling the effects of variations on the inter-residue non-covalent interaction network. The main difference with SDM is the presence, within mCSM-PPI2, of an interface analysis. DUET [81] combines two complementary approaches (mCSM and SDM) in an optimized predictor using support vector machines (SVM). mCSM-Membrane [82] is a machine-learning approach that uses graph-based structural signatures to test predictive models and analyze the impacts of mutations on the stability of membrane proteins and their possible association with diseases. mCSM-AB [83] is a machine-learning approach based on graph-based structural signatures that predict the change in antibody–antigen affinity upon introducing a single mutation. Finally, DynaMut2 [84] combines graph-based signatures with normal mode dynamics to predict the impact of actions on protein dynamics and stability resulting from vibrational entropy changes.

The described prediction tools are routinely used for pathogenicity prediction or rationalization of clinical evidence, and examples of specific applications to the field of PC are also reported. The web server for predictions of protein stability changes upon mutations (MUpro) has been employed to develop an index of responsiveness to pharmacological chaperones of variants of α -galactosidase in Fabry disease, scoring an 81% accuracy [85] for the PC Migalstat (now marketed as Galafold) vs. variants of α -galactosidase. Again for Fabry disease, PolyPhen-2 and SIFT have been used to support experimental data in the clinical evaluation of p.Asp313Tyr variant. The pathogenicity of such mutation has been debated since its first description in 1994 [86]. Recent studies converge to propose such a variant conducive to FD but with a milder phenotype and with later onset of symptoms [87–90]; predictions converge as well, reporting the variant as probably damaging (PolyPhen-2) or damaging (SIFT) [90].

5. Protein Instability Prediction by MD Simulation

The previously described methods are trained on datasets of protein sequences and structures and return scalar values, making them very effective in summarizing the effect

of mutations and processing large amounts of proteins/mutations. However, biological activity relies not only on chemical factors but also on dynamics, regulating, e.g., active site adaptability, loop flexibility, domain swapping, and more, to investigate which time-dependent representation of wild-type/mutant proteins is required. Although experimental techniques may investigate protein dynamics (e.g., nuclear magnetic resonance relaxation), it is much more common and less expensive to adopt a computational biology approach to simulate molecular dynamics. In particular, molecular dynamics (MD) simulations based on the integration of Newtonian equations of motion, boosted by modern GPU platforms, can simulate the dynamics of a protein at an atomic resolution for a time range of up to milliseconds. Different timescale dynamics can be explored, from aromatic ring flipping to domain swap. The most used suites for MD simulations are GROMACS [91,92], Amber [93], NAMD [94], Desmond [95], and Tinker [96].

MD simulations start from a three-dimensional structure but in the case of *ab initio* folding. Experimental structures from the Protein Data Bank [97] are the ideal starting point, provided that the resolution is fair. Unlike the prediction methods already illustrated, prosthetic groups or post-translational modifications in MD simulation cannot be easily neglected. Their inclusion is a non-trivial task and must be evaluated. Similarly, missing ligands can be omitted or added by, i.e., docking (see “Virtual screening” section). MD simulations are used either to investigate local perturbation or overall folding instability introduced by mutations. The perturbation analysis is the most common: time evolution of the structure (the so-called trajectory) of the native protein and its mutants are compared either visually or by plot analysis of parameters like root mean square fluctuations, H-bond pattern, secondary structure conservation/disruption, etc. The comparison allows for identifying and quantifying structural deviances anywhere in the structure, including secondary structure, active sites, interfaces, etc. It must be noted that since the MD simulation of the mutant protein starts from a 3D structure, the simulation makes sense only if there is evidence that the mutant reaches a folded state, which is, as discussed before in the text, not straightforward, even for a missense mutation. Strong evidence is, of course, the presence of the mutant as PDB experimental structure, but also residual biological activity (e.g., enzymatic) can be evaluated, as well as the convergence of stability prediction by the methods mentioned above (see section “Predicting the pathogenicity of variants and the effect on protein stability”). When the mutant protein is unavailable as a PDB structure, the same can be modeled *in silico* starting from the native structure. Missense mutations are easily modeled with PyMol [98] molecular viewer. Its mutagenesis tool provides a point-and-click interface to substitute amino acids in the 3D view of the molecule. The mutated sidechain is modeled *in situ* according to a library of energetically favorite conformers; the conformer with the best score in terms of fewer clashes/favorable interactions is proposed; the user can also explore other solutions, e.g., by modifying the χ angles of sidechains.

Mutation-induced perturbation can be quantified more accurately by free energy calculation in alchemical transformation [99,100]. Here, the comparison of two MD trajectories (native/mutated) is replaced by a single trajectory where the interested amino acid sidechain structure is morphed into the new one along the MD run (hence the name alchemical). Energy evaluation of the resulting transformation trajectory with the Bennett acceptance ratio method (BAR) [101,102] or accelerated weighted histogram (AWH) [100,103] allows for free energy perturbation to be evaluated.

When a mutation leads to misfolding, MD techniques may still help investigate the deleterious effects of the unfolding simulations. Indeed, experiments have demonstrated that unfolding MD simulations also shed light on folding, following the principle of microscopic reversibility, which states that folding and unfolding events are the same under the same conditions [104]. Thermal unfolding MD simulations obey the rules of Arrhenius. An increase in temperature does not modify the unfolding pathway but only its rate: the unfolding timescale is shortened as temperature increases. At the same time, the overall behavior and order of events are conserved [105]. In practice, the native and mutant

proteins undergo multiple simulations at increasing temperatures, e.g., 300 K, 500 K, 700 K, and more, to achieve different unfolding rates and extents (see Figure 3).

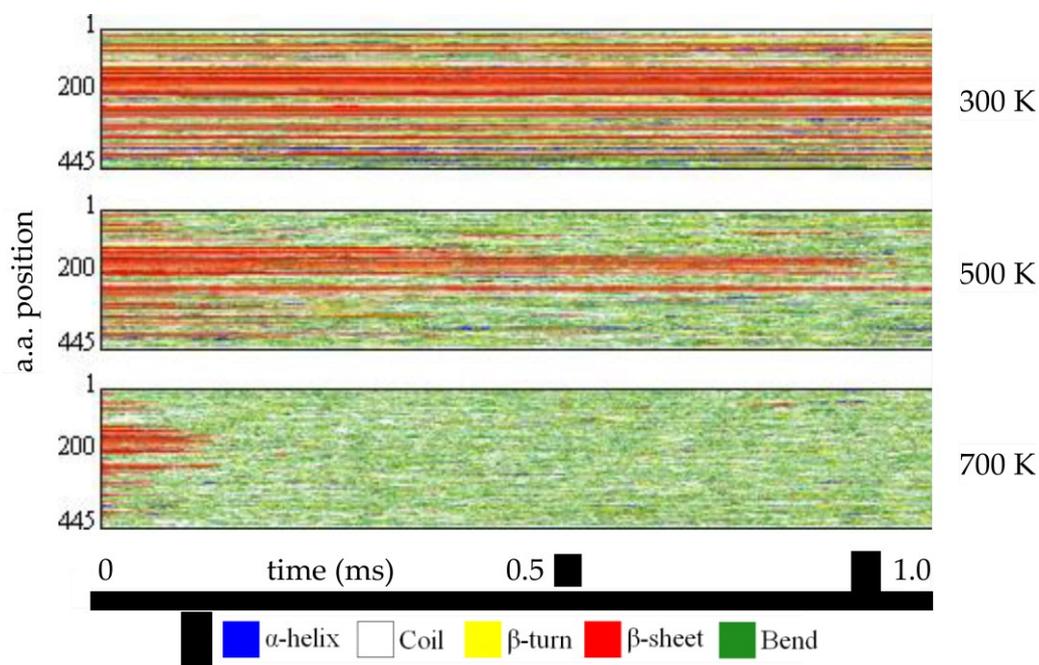


Figure 3. Example of unfolding MDs for wild-type HGD protomer from *H. sapiens* (Bernini A., unpublished data). The protein hydrated in a water box has been simulated at room temperature (300 K) and in increasing unfolding conditions (500 K and 700 K) for 1 ms with GROMACS and the amber force field. For each trajectory, the secondary structure content along the protein sequence (ordinate) has been evaluated as a function of time (abscissa) and plotted in different colors (see legend). A different rate of unfolding for the β -sheets is apparent by comparing the pictures.

The trajectories are then compared pairwise (native/mutant) with the analysis tools already described to highlight the structural moieties that unfold early/late. The influence of mutation can be evaluated. Although high-temperature unfolding MD simulation conditions appear unnatural, they benchmark well with NMR experimental data obtained at 298 K over the millisecond timescale [106].

Unfolding simulations have also performed well by replacing high temperatures with a denaturing environment like urea solution [107].

6. Pocket Prediction Tools

The topography of a protein structure is sculpted with many surface pockets and crevices, internal cavities, and channels. For example, ligands can bind to enzymes, and gas molecules can be routed through such microenvironments. Such microenvironments are best mapped by direct observation of a protein structure complexed with natural or engineered ligands (e.g., substrates, inhibitors), as found in PDB. Considering the reasons outlined in the introductory paragraphs, the trend over the past few years has been to identify PCs that bind to alternative sites of proteins rather than competing with the known ligand [20]. This can be experimentally achieved by, e.g., the use of molecular probes like in the multiple solvent crystal structure (MSCS) X-ray diffraction technique [108–110] or fragment-based NMR [111,112]. Alternative binding sites can also be discovered using bioinformatics tools mapping the surface/volume of a protein structure, allowing the search to be expanded to modeled proteins. This means virtually all the Uniprot (<https://www.uniprot.org/> (accessed on 10 January 2023)) [113] entries missing 3D experimental structure, now provided with the AlphaFold [114] structure prediction (536,000 reviewed entries counted on 4 January 2022, limited to the Swiss-prot section and excluding those

reporting experimental 3D structure). The tools for pocket/cavities discovery may rely on geometrical analysis, machine learning, or template-based approaches and are offered as web-based services (see Table 3). CASTp [115] is a metadatabase of topographies of PDB structures computed by grid-based geometry, but user PDB files can also be uploaded for computation. Results are displayed in a clean graphical viewer with pockets as bobble objects. 3DLigandSite [116] is a prediction tool using a template library of known binding sites; if the user cannot provide the 3D structure of the protein, the service retrieves it from AlphaFold [114] database or eventually models it with Phyre2 [117]. Prediction of protein-ligand binding residues is provided as a part of the intFOLD [118] modeling server, which uses the FunFOLD template-based algorithm; the online service starts from a protein sequence, and no custom PDB files are allowed. DeepSite [119] is a neural network-based predictor that works with PDB codes and user-supplied structures; results can be conveniently downloaded as comma-separated values (.csv) files and cavities as Gaussian format (.cube), suitable for more detailed analysis in, e.g., PyMol by representation as isomesh surface. COACH-D [120] is a ligand binding site predictor based on a support vector machine and refined by docking. The user can optionally upload a ligand structure that will be docked to the predicted pockets with Autodock Vina (discussed later), giving more targeted results. For modeling, proteins can be uploaded as PDB structure files or sequences in FASTA format.

Table 3. Web tools for predicting protein pockets/cavities with the corresponding website and method.

Name	URL	Method
CASTp	http://sts.bioe.uic.edu/castp (accessed on 10 January 2023)	grid-based geometry
3DLigandSite	https://www.wass-michaelislab.org/3dligandsite (accessed on 10 January 2023)	template-based
IntFOLD	https://www.reading.ac.uk/bioinf/IntFOLD (accessed on 10 January 2023)	template-based
DeepSite	https://playmolecule.com/deepsite (accessed on 10 January 2023)	template-based, neural networks
COACH-D	https://yanglab.nankai.edu.cn/COACH-D (accessed on 10 January 2023)	consensus, SVM
PrankWeb	http://prankweb.cz (accessed on 10 January 2023)	template-free, random forest

Unlike the previous methods, Prankweb [121] (a web frontend to P2Rank) offers a template-free machine-learning method based on the prediction of surface ligandability, i.e., the convergence of favorable chemical interaction points on a solvent-accessible protein surface.

The prediction of novel binding pockets on mutated proteins causing conformational diseases offers a valuable solution also when PC is already available but belongs to the competitive inhibitors class (see Section 2). The effectiveness of this kind of PC relies upon a delicate equilibrium with the substrate, which can be avoided by substituting the competitive PC with a stabilizer binding an alternative site, e.g., allosteric (see the Tafamidis/TTR example) or, even better, devoid of biological function. Thus, pocket prediction offers a solution for the development of PC of the second generation. For example, in Fabry disease research, such an approach allowed the discovery of a different druggable pocket located at the opposite side of the active site in α -galactosidase [122]. Such pocket is capable of binding 2,6-dithiopurine, and the binding stabilizes the enzyme in vitro and, noticeably, rescues the p.Ala230Thr variant that is not responsive to monotherapy with Migalstat (the competitive inhibitor PC for α -galactosidase, see Table 1). A similar effort has been carried out to overcome the drawbacks of inhibitory chaperones in Gaucher disease: a series of pyrrolopyrazines [123] have been demonstrated to bind to a new pocket at a dimer interface and induce dimerization.

7. Exploiting Transient Pockets for PC Binding

The methods illustrated in the previous paragraph allow for alternative pockets to be identified on a protein surface, but, despite the different approaches, they all share the limit of computing on a single structure, usually a conformation of minimum energy. As a result, the protein's dynamics are neglected, a limit already highlighted for stability prediction methods. While molecular dynamics trajectories are usually explored focusing on binding sites and secondary structures, protein surfaces deserve more attention in the search for alternative binding sites, as they may "hide" transient pockets.

The term "transient pocket" refers to a crevice opening at the protein surface due to side chains or backbone dynamics which is not detectable in the energy-minimized or averaged structure. When discussed in conjunction with drug binding, such pockets are also referred to as "cryptic" in literature [124]. Transient or cryptic pockets are valuable and provide a promising alternative to classical binding sites for drug development or repositioning.

The concept of protein transient pocket was first exploited as a target for disruptor of protein-protein interaction (PPI) since such interfaces are usually flat and strategies for drug design aiming at "classical" targets, e.g., G-protein-coupled receptors and enzymes, do not apply [125]. PPI investigations have shown that usually, in the center of the contact interface, a small subset of the residues complex contributes most of the free energy of binding. These hotspots have been exploited as receptor sites. Furthermore, once the PPI disruptor binds to the hotspot, the flat surface forms a groove, as has been observed in different protein-protein interaction systems, such as IL-2, Bcl-Xl, and HDM2 [125]. Moreover, molecular dynamics (MD) has shown that in the absence of PPI disruptors, proteins' transient pockets open in less than one nanosecond [126].

The nature of the transient pockets has been studied with many alternative MD methods (NMA, CONCOORD, and tCONCOORD) [127] and different solvents (water and methanol) [128]. The result was that the number of transient pockets grew with methanol compared to those found in water, suggesting a lipophilic nature of the cavities. Moreover, it has been suggested that side chain movements are insufficient for a transient pocket to open; backbone dynamics are also necessary.

In the last decade, transient pockets have been evaluated for the re-gain of activity of mutated proteins, as for the Y220C mutant of p53 tumor-suppression protein [129]. A first MD simulation and docking study identified a transient pocket opening transiently at Loop 1/ β -strand 3 [130]. In successive research [129], several molecules were found binding at a transient pocket site. Such crevice has its maximum depth at the mutation site (Y220C) (Figure 4). A similar study has been carried out on the small chemokine CXCL12 (responsible for tumor progression and proliferation), where a transient pocket exploitable for binding small molecules has been observed both in MD simulation and in vitro by NMR experiment [131].

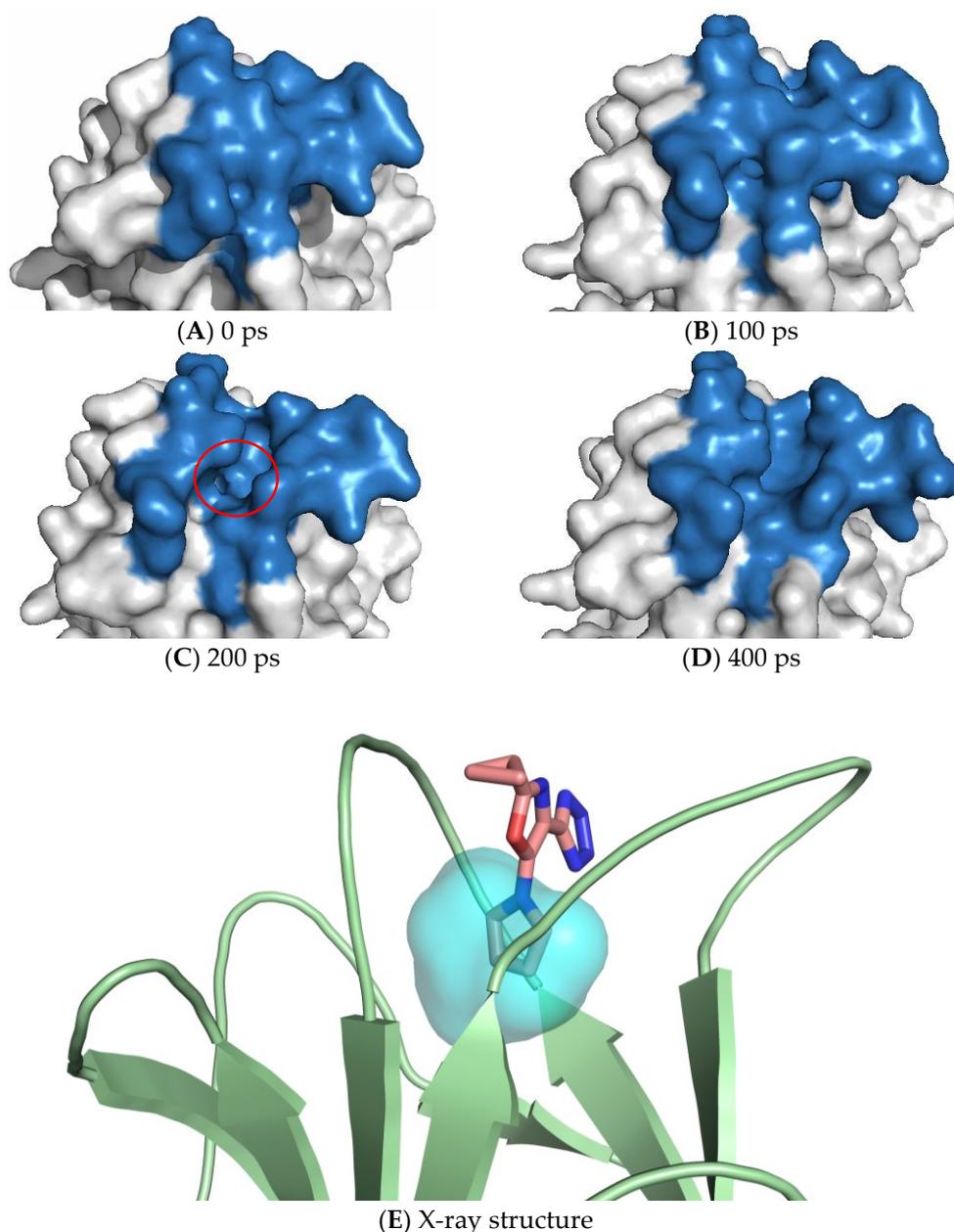


Figure 4. (A–D) Snapshots from an MD simulation showing the gradual opening/closing of a transient pocket on the surface region (blue) surrounding the Y220C mutation of the protein p53 (white). The red circle spots the transient pocket in the maximum depth conformation. (E) The transient pocket identified in the MD superposed to that occupied by the pyrrolic moiety of PhiKan7099 small molecule ligand (in pink) [132] in the experimental structure from PDB ID 5AOK (accessed on 10 January 2023).

Furthermore, molecular dynamics simulation of the HIV-1 protease has shown a transient pocket with a lifetime longer than 1 μ s that was used as the target for docking studies aiming to discover a new anti-AIDS drug [133].

The search for transient pockets available for small molecules is increasingly important indeed. For this reason, it requires effective programs to identify the latter. Programs for transient pocket prediction are usually MD trajectory-based, such as EPOS [126] and Mdpocket [134]. EPOS searches for transient pockets by analyzing a sequence of MD snapshots using the PASS pocket detection algorithm [135]. Similarly, Mdpocket analyzes a

set of structural snapshots sampled along the molecular dynamics trajectory by exploiting the fpocket program based on the Voronoi tessellation [134,136].

8. Druggability of Pockets

The approaches to pocket discovery described above open the way to a fundamental question: what causes a binding site to be a binding site? Indeed, in a drug-repositioning approach, finding alternative pockets is not enough; they must also be suitable for drug binding, i.e., “druggable”. Indeed, by looking at a binding pocket, the shape is the most evident trait, but the chemical properties cause it to be “binding”. The ability to efficiently bind a ligand is estimated from the possible number and types of non-covalent interactions that can be established, the so-called ligandability or druggability in medicinal chemistry research. A druggable pocket should reflect the typical features of a drug molecule, which are usually coded by Lipinski’s “Rule of Five” (Ro5) [137]. Ro5 describes the physicochemical characteristic required to be orally active for small molecules. According to the Ro5, a compound should possess the following properties to be active: (i) a molecular mass less than 500 Da; (ii) high lipophilicity: the octanol–water partition coefficient (LogP) should be less than 5; (iii) at most, five hydrogen bond donors; and iv) less than 10 hydrogen bond acceptors. All indicators are multiples of five indeed. A critical discussion on the effectiveness of the Ro5 covering two decades of research is reported in [138].

Predicting a protein pocket’s druggability is one of the most critical steps in drug discovery and development, now supported by many computational tools (Table 4). PockDrug [139] is a server relying on multiple pocket estimation methods allowing the prediction of pocket druggability of both apo and holo proteins with good accuracy. PockDrug combines four pocket estimation methods to evaluate the best pocket set from the ‘NonRedundant dataset of Druggable and Less Druggable binding sites’ (NRDLD) [140]. After model validation, PockDrug outputs druggability probability corresponding to the average of the seven best models and its standard deviation. Fpocket [136] is based on the geometric α -sphere theory; the Fpocket server consists of three software packages: (1) pocket identification using the concept of α -spheres (fpocket package); (2) pocket tracking along molecular dynamics trajectory (mdpocket package); and (3) collection of homologous structures using the pocket tracking (hpocket package). Through these three steps, the program allows the identification of both allosteric sites and transient pockets. DoGSiteScorer [141] is another server to analyze pockets’ geometric and physicochemical properties and predict their druggability; identifying possible pockets is performed using the protein’s heavy atom coordinates and calculating a density threshold. Based on a supporting vector machine (SVM), the algorithm proceeds for subpockets identification that are then blended into pockets. Protein–ligand interaction clusters (PLIC) [142] are a metadatabase that provides binding sites from the Protein Data Bank clustered by similarity using Markov clustering (MCL) [143] algorithm; by using computational tools, such as SPACE [144], fPocket [136], and Autodock [145], it also calculates various attributes of the binding sites, such as binding energy, shape, polarity, hydrophobicity, etc. CavityPlus [146] is a web server for pocket detection and functional analysis. The Cavity subroutine identifies cavities on protein surfaces using the NRDLD [140] dataset to train and validate the model. For each of them, CavityPlus provides pharmacophore modeling, allosteric site identification, and covalent ligandability prediction through the subroutines CavPharmer, CorrSite, and CovCys, respectively. PharmMapper Server [147,148] is a web server that, contrary to the others, starts from a small molecule probe and identifies potential targets using a pharmacophore mapping approach. PharmMapper relies on an in-house pharmacophore database obtained from the combination of other small molecule databases. In PharmMapper, target proteins with the highest fit scores between corresponding pharmacophore models and query compounds are reported as potential targets.

Table 4. Freely available online tools for the prediction of protein pockets' druggability and their web addresses. Methods applied are described in the text, along with references.

Name	URL
PockDrug	http://pockdrug.rpbs.univ-paris-diderot.fr/ (accessed on 10 January 2023)
Fpocket	https://fpocket.sourceforge.net/ (accessed on 10 January 2023)
DoGSiteScorer	https://proteins.plus/ (accessed on 10 January 2023)
CavityPlus	http://www.pkumdl.cn:8000/cavityplus/ (accessed on 10 January 2023)
PharmMapper	http://www.lilab-ecust.cn/pharmmapper/ (accessed on 10 January 2023)
PLIC	http://proline.biochem.iisc.ernet.in/PLIC/ (accessed on 10 January 2023)

9. Virtual Screening of Compound Libraries in Search of PCs

Once a protein mutation and alternative binding site have been positively evaluated as a target for PC-mediated rescue, the hit discovery phase starts. Again, computational methods can be exploited to reduce time and cost in a workflow targeted to rare diseases; the list of compounds to be evaluated by experimental assay is slimmed down through simulated molecular docking of large compound databases [149]

A critical analysis of more than 400 studies involving virtual screening of small molecule libraries showed an overall hit rate ranging from 1% to 40% [150], comparable and somewhat superior to high throughput screening [151] both in terms of hit and screened compounds. A significant contribution to virtual screening comes from freely available databases of small molecule ligands (Table 5). Among the most popular, reporting drug-like, bioactive, and synthesizable compounds are ZINC [152], PubChem [153], DrugBank [154], ChEMBL [155], e-Drug3D [156], SuperDRUG2 [157], BindingDB [158], HMDB [159], and LIGAND [160]. The size of such molecular datasets, counting millions of compounds, has been expanded to billions of entries by enumerating synthetically feasible molecules (REAL: REAdily Available for synthesis Library [161], 5.5B compounds at the time of writing) or by enumerating molecules in virtual chemical space (as in the Chemical Universe Database GDB-17 [162], reporting 166B compounds).

Table 5. Free accessible databases of small molecule ligands and their corresponding web address.

Name	URL
ZINC20	https://zinc20.docking.org/ (accessed on 10 January 2023)
PubChem	https://pubchem.ncbi.nlm.nih.gov/ (accessed on 10 January 2023)
DrugBank	https://go.drugbank.com/ (accessed on 10 January 2023)
ChEMBL	https://www.ebi.ac.uk/chembl/ (accessed on 10 January 2023)
e-Drug3D	https://chemoinf.ipmc.cnrs.fr/MOLDB/index.php (accessed on 10 January 2023)
SuperDRUG2	http://bioinf.charite.de/superdrug (accessed on 10 January 2023)
BindingDB	https://www.bindingdb.org/bind/index.jsp (accessed on 10 January 2023)
HMDB	https://hmdb.ca/ (accessed on 10 January 2023)
Ligand	https://www.genome.jp/kegg/compound/ (accessed on 10 January 2023)
REAL	https://enamine.net/compound-collections/real-compounds (accessed on 10 January 2023)
GDB17	https://gdb.unibe.ch/downloads/ (accessed on 10 January 2023)

Molecular structures are usually available as 3D-coordinate file formats, such as sdf, mol2, and pdbqt (an extended format for pdb, including extra columns for atom charge and solvation energy). Other formats are becoming popular, such as the SMILES [163], which encodes the molecule topology into a lightweight ASCII string at the cost of losing conformation; on the contrary db2 format contains not one but multiple pre-calculated conformations [152] at the expense of larger file size. The protein molecule is the “target” of the screening and is usually provided in pdb format; the experimental or modeled structure or a snapshot from an MD simulation trajectory can be employed. Screening can be replicated over multiple trajectory snapshots to expand the target's conformational sampling [164,165]. Usually, only the volume of the target containing the candidate pocket is screened, called the “docking box,” because of its simple shape. A box shape is used

because of the easiness of mathematical representation in the pdb format, with three coordinates for the origin and three for the side lengths.

Virtual screening can be performed with molecular docking simulation software that evaluates the ligand–target interaction energy according to a scoring function based on atom partial charge, desolvation energy, and others. The docking is flexible both on the ligand and target sides. Ligands are assigned rotatable bonds at system set-up, the same for protein but limited to amino acidic sidechains in the docking box. Among the freely available docking software, many belong to the Autodock family: Autodock 4 [166,167], Autodock Vina [168,169], AutoDockFR [170], FRED [171], EnzyDock [172], ICM [173], FlexX [174], Glide [175], GOLD [176], Mdock [177], and MOE [178]. Most of the mentioned software programs have also been optimized for parallel execution [179] and are readily available on high-performance computers with high scalability, allowing larger datasets to be screened.

The choice of docking software is connected mainly with the speed and accuracy of the scoring function, and performance comparison is largely covered in the literature [180–183]. Ease of usage is nowadays a standard, thanks to many graphical interfaces such as AutoDockTools [166], Raccoon v1.0 (AutoDock4) [184], and Raccoon v2.0 (AutoDock Vina) [184]. The typical output of a virtual screening with docking software is a list of compounds ranked by the score. The top-scorer is not necessarily the best hit [185]; it is then a good practice to take a top group of compounds and re-rank according to other scoring functions, molecular dynamics simulation-derived free energy of binding (e.g., MM/PBSA and MM/GBSA [186]), and, of course, experimental validation.

Molecular docking and MD simulation studies in the search for PC against conformational diseases have become more and more relevant in the recent literature. As an example, many can be found about the discovery of PC for β -glucocerebrosidase, the target protein in Gaucher disease [187–191]. Similar examples are reported for PC development rescuing β -hexosaminidase A (associated with Tay-Sachs disease) [192] and N-acetylgalactosamine-6-sulfate sulfatase (responsible for mucopolysaccharidosis IV A) [193].

Recently, machine learning [194–197] and deep learning by artificial intelligence [198–200] approaches have succeeded in molecular docking, enabling the virtual screening of libraries of up to billions of compounds. The first experimental validations of deep docking are also encouraging [201]. A breakthrough of AI in virtual screening similar to AlphaFold [114] on structural biology is then envisaged and covered in Section 10.

10. The Impact of Artificial Intelligence

The re-emergence of artificial intelligence (AI) is having an impact on the drug-discovery process, e.g., by expanding the chemical space of virtual screening but also on de novo design of active compounds. Similar to drug-repositioning intent but with opposite approach, the advantage of such AI application in PC development for rare diseases is that they allow for scaffold hopping [202], which may help establish a competitive patent position [203], and are here described in their latest implementations in virtual docking and de novo design.

Despite its benefits, virtual docking ultimately lags behind the rapid expansion of chemical databases which already exceed billions of records when going to virtual chemical space (e.g., the Chemical Universe Database GDB-17 [162] enumerates 166 billion organic small molecules).

In order to address this challenge, AI is being introduced in the process and many platforms are already available. An example is the deep-learning platform Deep Docking (DD) [204]: small portions of ultra-large databases are used to train quantitative structure-activity relationship models that predict remaining entries' scores, accelerating virtual screening 50 times. DD can now also take advantage of a convenient graphical user interface (DD-GUI) [205] for quick setups and outcome analyses of large-scale virtual screens. A similar approach of chemical space reduction, based on Bayesian optimization techniques, has been proposed to speed up the exploration of large molecular libraries by

developing a surrogate model architecture [206]: in order to exclude the least promising compounds from evaluation, a surrogate structure–property relationship model can be created using the predicted affinities of a subset of the library. Inspired by natural processes, the ant colony optimization algorithm (ACO) is at the base of an artificial intelligence driven docking algorithm recently implemented in VirtualFlow [207], the VirtualFlow Ants [208]. The software showed almost linear scalability to 128,000 CPUs on a test library of a total of 1.44 billion compounds using KEAP1 protein as target. As an added value, it allows multiple re-scoring and on-the-fly conversion of structure formats.

PyRMD [198] is a fully automated AI-powered ligand-based virtual screening tool that implements in Python the random matrix discriminant (RMD) method, capable of predicting whether a given small molecule could bind to a certain target. It requires only a set of active ligands on a specific target to train the algorithm, making it entirely ligand-based. A synthon-based approach, working in tight conjunction with REAL libraries [161], has been proposed in V-SYNTHES [209]. As the name suggests, the method is based on synthons; the synthons are those simple, charged chemical species that are formed during retrosynthetic analysis after imaginatively breaking down the target molecule.

While the goal of AI in the methods described above is to allow structure-based virtual screening to exploit compound libraries that are growing larger and larger, a different approach of “de novo” design is resurging thanks to machine intelligence development [210]. As the name implies, de novo design is the process of designing from scratch new chemical entities with desired (pharmacological) properties. Ideally, machine intelligence may jointly perform the three steps involved in the de novo design of molecules: generation, scoring, and optimization.

In this view, the generative artificial intelligence approach [211] offers an AI model for the design of drug-like molecules. Such a deep-learning method is capable of “capturing” chemical determinants from a training set of drugs already known as active against a target protein and designing new chemical entities; it has been successfully applied to synthesizing novel active agonists of peroxisome proliferator-activated receptors. Similarly, generative adversarial networks have been proposed [212] for the de novo design of molecules from chEMBL [155] database. Reinforcement learning (RL) is at the base of ReLeaSE (Reinforcement Learning for Structural Evolution) strategy [213], targeted at generating chemical libraries of novel compounds optimized for either a single desired property or multiple properties. An up-to-date review of such a growing area of research can be found in [214].

11. Conclusions and Future Directions

Conformational diseases result from misfolding of proteins, often due to missense pathological variants causing the loosening (or sometimes the strengthening) of the three-dimensional structure or quaternary assembly of the macromolecule. It has been shown that drug-like molecules may bind and stabilize the structure of poorly folded proteins, leading to the rescue of structure and activity, hence the name pharmacological chaperones. Furthermore, protein conformational diseases often superpose with rare genetic disorders, requiring strategies for lower drug development costs. The present review shows how the computational structural biology approach finds a convenient application in pharmacological chaperone development. We described a general workflow (Figure 5) that lines up several *in silico* tools for protein target discovery and pathogenicity prediction of variants (protein stability predictors and MD simulations) with those already used in rational drug design (pocket discovery, pocket druggability and virtual screening).

The next challenge for such a strategy will be replacing knowledge-based predictions with deep learning, which has already overperformed classical methods in structural biology.

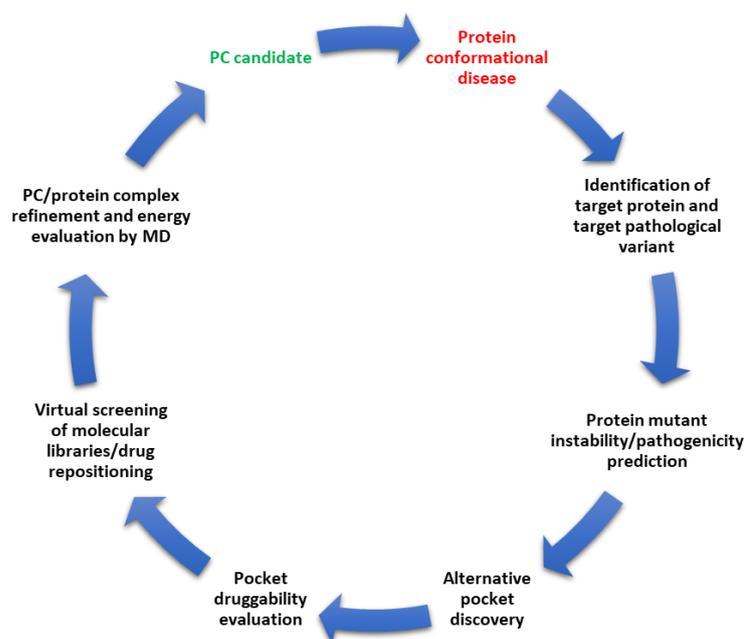


Figure 5. A general workflow for the application of computational structural biology to the development of pharmacological chaperones.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: A.B. thanks Monica Tiezzi for the helpful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Reynaud, E. Protein Misfolding and Degenerative Diseases. *Nat. Educ.* **2010**, *3*, 28.
- Soto, C. Protein Misfolding and Disease; Protein Refolding and Therapy. *FEBS Lett.* **2001**, *498*, 204–207. [[CrossRef](#)] [[PubMed](#)]
- Prabantu, V.M.; Naveenkumar, N.; Srinivasan, N. Influence of Disease-Causing Mutations on Protein Structural Networks. *Front. Mol. Biosci.* **2021**, *7*, 492. [[CrossRef](#)] [[PubMed](#)]
- Valastyan, J.S.; Lindquist, S. Mechanisms of Protein-Folding Diseases at a Glance. *Dis. Model. Mech.* **2014**, *7*, 9–14. [[CrossRef](#)]
- Bross, P.; Andresen, B.S.; Corydon, T.J.; Gregersen, N. Protein Misfolding and Degradation in Genetic Disease. *Hum. Mutat.* **1999**, *14*, 186. [[CrossRef](#)]
- Miosge, L.A.; Field, M.A.; Sontani, Y.; Cho, V.; Johnson, S.; Palkova, A.; Balakishnan, B.; Liang, R.; Zhang, Y.; Lyon, S.; et al. Comparison of Predicted and Actual Consequences of Missense Mutations. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5189–E5198. [[CrossRef](#)]
- Andrews, T.D.; Sjollem, G.; Goodnow, C.C. Understanding the Immunological Impact of the Human Mutation Explosion. *Trends Immunol.* **2013**, *34*, 99–106. [[CrossRef](#)]
- Peng, Y.; Alexov, E. Protein Conformational Disease: Visit the Facts at a Glance. *eLS* **2017**. [[CrossRef](#)]
- Conde-Giménez, M.; Galano-Frutos, J.J.; Galiana-Cameo, M.; Mahía, A.; Victor, B.L.; Salillas, S.; Velázquez-Campoy, A.; Brito, R.M.M.; Gálvez, J.A.; Díaz-de-Villegas, M.D.; et al. Alchemical Design of Pharmacological Chaperones with Higher Affinity for Phenylalanine Hydroxylase. *Int. J. Mol. Sci.* **2022**, *23*, 4502. [[CrossRef](#)]
- Chen, A.; Pan, Y.; Chen, J. Clinical, Genetic, and Experimental Research of Hyperphenylalaninemia. *Front. Genet.* **2022**, *13*, 1051153. [[CrossRef](#)] [[PubMed](#)]
- Grasso, D.; Geminiani, M.; Galderisi, S.; Iacomelli, G.; Peruzzi, L.; Marzocchi, B.; Santucci, A.; Bernini, A. Untargeted NMR Metabolomics Reveals Alternative Biomarkers and Pathways in Alkaptonuria. *Int. J. Mol. Sci.* **2022**, *23*, 5805. [[CrossRef](#)] [[PubMed](#)]
- Zatkova, A.; Sedlackova, T.; Radvansky, J.; Polakova, H.; Nemethova, M.; Aquaron, R.; Dursun, I.; Usher, J.L.; Kadasi, L. Identification of 11 Novel Homogentisate 1,2 Dioxygenase Variants in Alkaptonuria Patients and Establishment of a Novel LOVD-Based HGD Mutation Database. *JIMD Rep.* **2012**, *4*, 55–65. [[CrossRef](#)]
- Echaniz-Laguna, A.; Cauquil, C.; Labeyrie, C.; Adams, D. Treating Hereditary Transthyretin Amyloidosis: Present & Future Challenges. *Rev. Neurol.* **2023**, *179*, 30–34. [[CrossRef](#)] [[PubMed](#)]

14. Shaimardanova, A.A.; Solovyeva, V.V.; Issa, S.S.; Rizvanov, A.A. Gene Therapy of Sphingolipid Metabolic Disorders. *Int. J. Mol. Sci.* **2023**, *24*, 3627. [[CrossRef](#)]
15. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. [[CrossRef](#)] [[PubMed](#)]
16. Bauer, N.C.; Doetsch, P.W.; Corbett, A.H. Mechanisms Regulating Protein Localization. *Traffic* **2015**, *16*, 1039–1061. [[CrossRef](#)] [[PubMed](#)]
17. Aymami, J.; Barril, X.; Rodríguez-Pascau, L.; Martinell, M. Pharmacological Chaperones for Enzyme Enhancement Therapy in Genetic Diseases. *Pharm. Pat. Anal.* **2013**, *2*, 109–124. [[CrossRef](#)] [[PubMed](#)]
18. Muntau, A.C.; Leandro, J.; Staudigl, M.; Mayer, F.; Gersting, S.W. Innovative Strategies to Treat Protein Misfolding in Inborn Errors of Metabolism: Pharmacological Chaperones and Proteostasis Regulators. *J. Inherit. Metab. Dis.* **2014**, *37*, 505–523. [[CrossRef](#)]
19. Ringe, D.; Petsko, G.A. What Are Pharmacological Chaperones and Why Are They Interesting? *J. Biol.* **2009**, *8*, 80. [[CrossRef](#)]
20. Parenti, G.; Moracci, M.; Fecarotta, S.; Andria, G. Pharmacological chaperone therapy for lysosomal storage diseases. *Future Med. Chem.* **2014**, *6*, 1031. [[CrossRef](#)]
21. Leeson, P.D.; Springthorpe, B. The Influence of Drug-like Concepts on Decision-Making in Medicinal Chemistry. *Nat. Rev. Drug. Discov.* **2007**, *6*, 881–890. [[CrossRef](#)] [[PubMed](#)]
22. Parenti, G.; Andria, G.; Valenzano, K.J. Pharmacological Chaperone Therapy: Preclinical Development, Clinical Translation, and Prospects for the Treatment of Lysosomal Storage Disorders. *Mol. Ther.* **2015**, *23*, 1138–1148. [[CrossRef](#)] [[PubMed](#)]
23. Asano, N.; Ishii, S.; Kizu, H.; Ikeda, K.; Yasuda, K.; Kato, A.; Martin, O.R.; Fan, J.Q. In Vitro Inhibition and Intracellular Enhancement of Lysosomal Alpha-Galactosidase A Activity in Fabry Lymphoblasts by 1-Deoxygalactonojirimycin and Its Derivatives. *Eur. J. Biochem.* **2000**, *267*, 4179–4186. [[CrossRef](#)]
24. Yam, G.H.-F.; Zuber, C.; Roth, J. A Synthetic Chaperone Corrects the Trafficking Defect and Disease Phenotype in a Protein Misfolding Disorder. *FASEB J.* **2005**, *19*, 12–18. [[CrossRef](#)]
25. Fan, J.Q.; Ishii, S.; Asano, N.; Suzuki, Y. Accelerated Transport and Maturation of Lysosomal Alpha-Galactosidase A in Fabry Lymphoblasts by an Enzyme Inhibitor. *Nat. Med.* **1999**, *5*, 112–115. [[CrossRef](#)] [[PubMed](#)]
26. Yam, G.H.F.; Bosshard, N.; Zuber, C.; Steinmann, B.; Roth, J. Pharmacological Chaperone Corrects Lysosomal Storage in Fabry Disease Caused by Trafficking-Incompetent Variants. *Am. J. Physiol. Cell Physiol.* **2006**, *290*, C1076–C1082. [[CrossRef](#)]
27. Lieberman, R.L.; Wustman, B.A.; Huertas, P.; Powe, A.C.; Pine, C.W.; Khanna, R.; Schlossmacher, M.G.; Ringe, D.; Petsko, G.A. Structure of Acid Beta-Glucosidase with Pharmacological Chaperone Provides Insight into Gaucher Disease. *Nat. Chem. Biol.* **2007**, *3*, 101–107. [[CrossRef](#)]
28. Khanna, R.; Benjamin, E.R.; Pellegrino, L.; Schilling, A.; Rigat, B.A.; Soska, R.; Nafar, H.; Ranes, B.E.; Feng, J.; Lun, Y.; et al. The Pharmacological Chaperone Isofagomine Increases Activity of the Gaucher Disease L444P Mutant Form of β -Glucosidase. *FEBS J.* **2010**, *277*, 1618. [[CrossRef](#)] [[PubMed](#)]
29. Staudigl, M.; Gersting, S.W.; Danecka, M.K.; Messing, D.D.; Woidy, M.; Pinkas, D.; Kemter, K.F.; Blau, N.; Muntau, A.C. The Interplay between Genotype, Metabolic State and Cofactor Treatment Governs Phenylalanine Hydroxylase Function and Drug Response. *Hum. Mol. Genet.* **2011**, *20*, 2628–2641. [[CrossRef](#)]
30. Michals-Matalon, K. Sapropterin Dihydrochloride, 6-R-L-Erythro-5,6,7,8-Tetrahydrobiopterin, in the Treatment of Phenylketonuria. *Expert Opin. Investig. Drugs* **2008**, *17*, 245–251. [[CrossRef](#)]
31. Pey, A.L.; Pérez, B.; Desviat, L.R.; Martínez, M.A.; Aguado, C.; Erlandsen, H.; Gámez, A.; Stevens, R.C.; Thórolfsson, M.; Ugarte, M.; et al. Mechanisms Underlying Responsiveness to Tetrahydrobiopterin in Mild Phenylketonuria Mutations. *Hum. Mutat.* **2004**, *24*, 388–399. [[CrossRef](#)]
32. Plumadore, E.; Lombardo, L.; Cabral, K.P. Pharmacotherapy Review: Emerging Treatment Modalities in Transthyretin Cardiac Amyloidosis. *Am. J. Health-Syst. Pharm.* **2022**, *79*, 52–62. [[CrossRef](#)]
33. Lamb, Y.N.; Deeks, E.D. Tafamidis: A Review in Transthyretin Amyloidosis with Polyneuropathy. *Drugs* **2019**, *79*, 863–874. [[CrossRef](#)] [[PubMed](#)]
34. Lamb, Y.N. Tafamidis: A Review in Transthyretin Amyloid Cardiomyopathy. *Am. J. Cardiovasc. Drugs* **2021**, *21*, 113–121. [[CrossRef](#)]
35. Bulawa, C.E.; Connelly, S.; DeVit, M.; Wang, L.; Weigel, C.; Fleming, J.A.; Packman, J.; Powers, E.T.; Wiseman, R.L.; Foss, T.R.; et al. Tafamidis, a Potent and Selective Transthyretin Kinetic Stabilizer That Inhibits the Amyloid Cascade. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9629–9634. [[CrossRef](#)]
36. Garcia-Pavia, P.; Domínguez, F.; Gonzalez-Lopez, E. Transthyretin Amyloid Cardiomyopathy. *Med. Clin.* **2021**, *156*, 126–134. [[CrossRef](#)]
37. Coelho, T.; Maia, L.F.; Da Silva, A.M.; Cruz, M.W.; Planté-Bordeneuve, V.; Lozeron, P.; Suhr, O.B.; Campistol, J.M.; Conceição, I.M.; Schmidt, H.H.J.; et al. Tafamidis for Transthyretin Familial Amyloid Polyneuropathy: A Randomized, Controlled Trial. *Neurology* **2012**, *79*, 785–792. [[CrossRef](#)] [[PubMed](#)]
38. Merlini, G.; Planté-Bordeneuve, V.; Judge, D.P.; Schmidt, H.; Obici, L.; Perlini, S.; Packman, J.; Tripp, T.; Grogan, D.R. Effects of Tafamidis on Transthyretin Stabilization and Clinical Outcomes in Patients with Non-Val30Met Transthyretin Amyloidosis. *J. Cardiovasc. Transl. Res.* **2013**, *6*, 1011–1020. [[CrossRef](#)]

39. Maurer, M.S.; Grogan, D.R.; Judge, D.P.; Mundayat, R.; Packman, J.; Lombardo, I.; Quyyumi, A.A.; Aarts, J.; Falk, R.H. Tafamidis in Transthyretin Amyloid Cardiomyopathy: Effects on Transthyretin Stabilization and Clinical Outcomes. *Circ. Heart Fail.* **2015**, *8*, 519–526. [[CrossRef](#)] [[PubMed](#)]
40. Kon, T.; Misumi, Y.; Nishijima, H.; Honda, M.; Suzuki, C.; Baba, M.; Inomata, Y.; Obayashi, K.; Ando, Y.; Tomiyama, M. Effects of Liver Transplantation and Tafamidis in Hereditary Transthyretin Amyloidosis Caused by Transthyretin Leu55Pro Mutation: A Case Report. *Amyloid* **2015**, *22*, 203–204. [[CrossRef](#)]
41. Scott, L.J. Tafamidis: A Review of Its Use in Familial Amyloid Polyneuropathy. *Drugs* **2014**, *74*, 1371–1378. [[CrossRef](#)]
42. Suhr, O.B.; Conceição, I.M.; Karayal, O.N.; Mandel, F.S.; Huertas, P.E.; Ericzon, B.G. Post Hoc Analysis of Nutritional Status in Patients with Transthyretin Familial Amyloid Polyneuropathy: Impact of Tafamidis. *Neurol. Ther.* **2014**, *3*, 101–112. [[CrossRef](#)] [[PubMed](#)]
43. Aflaki, E.; Borger, D.K.; Moaven, N.; Stubblefield, B.K.; Rogers, S.A.; Patnaik, S.; Schoenen, F.J.; Westbroek, W.; Zheng, W.; Sullivan, P.; et al. A New Glucocerebrosidase Chaperone Reduces α -Synuclein and Glycolipid Levels in iPSC-Derived Dopaminergic Neurons from Patients with Gaucher Disease and Parkinsonism. *J. Neurosci.* **2016**, *36*, 7441–7452. [[CrossRef](#)]
44. Suzuki, Y.; Ichinomiya, S.; Kurosawa, M.; Matsuda, J.; Ogawa, S.; Iida, M.; Kubo, T.; Tabe, M.; Itoh, M.; Higaki, K.; et al. Therapeutic Chaperone Effect of N-Octyl 4-Epi- β -Valienamine on Murine G(M1)-Gangliosidosis. *Mol. Genet. Metab.* **2012**, *106*, 92–98. [[CrossRef](#)]
45. Siriwardena, A.; Sonawane, D.P.; Bande, O.P.; Markad, P.R.; Yonekawa, S.; Tropak, M.B.; Ghosh, S.; Chopade, B.A.; Mahuran, D.J.; Dhavale, D.D. Synthesis of 1,5-Dideoxy-1,5-Iminoribitol C-Glycosides through a Nitron-Olefin Cycloaddition Domino Strategy: Identification of Pharmacological Chaperones of Mutant Human Lysosomal β -Galactosidase. *J. Org. Chem.* **2014**, *79*, 4398–4404. [[CrossRef](#)] [[PubMed](#)]
46. Pan, X.; Taherzadeh, M.; Bose, P.; Heon-Roberts, R.; Nguyen, A.L.A.; Xu, T.; Pará, C.; Yamanaka, Y.; Priestman, D.A.; Platt, F.M.; et al. Glucosamine Amends CNS Pathology in Mucopolysaccharidosis IIIC Mouse Expressing Misfolded HGSNAT. *J. Exp. Med.* **2022**, *219*, e20211860. [[CrossRef](#)]
47. Dawson, G.; Schroeder, C.; Dawson, P.E. Palmitoyl:Protein Thioesterase (PPT1) Inhibitors Can Act as Pharmacological Chaperones in Infantile Batten Disease. *Biochem. Biophys. Res. Commun.* **2010**, *395*, 66–69. [[CrossRef](#)]
48. Bier, D.; Bartel, M.; Sies, K.; Halbach, S.; Higuchi, Y.; Haranosono, Y.; Brummer, T.; Kato, N.; Ottmann, C. Small-Molecule Stabilization of the 14-3-3/Gab2 Protein-Protein Interaction (PPI) Interface. *ChemMedChem* **2016**, *11*, 911–918. [[CrossRef](#)]
49. Inglese, J.; Shamu, C.E.; Guy, R.K. Reporting Data from High-Throughput Screening of Small-Molecule Libraries. *Nat. Chem. Biol.* **2007**, *3*, 438–441. [[CrossRef](#)] [[PubMed](#)]
50. Langedijk, J.; Mantel-Teeuwisse, A.K.; Slijkerman, D.S.; Schutjens, M.H.D.B. Drug Repositioning and Repurposing: Terminology and Definitions in Literature. *Drug Discov. Today* **2015**, *20*, 1027–1034. [[CrossRef](#)] [[PubMed](#)]
51. Croset, S. Drug Repositioning and Indication Discovery Using Description Logics. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2014. [[CrossRef](#)]
52. Khanapure, A.; Chuki, P.; De Sousa, A. Drug Repositioning: Old Drugs For New Indications. *Indian J. Appl. Res.* **2011**, *4*, 462–466. [[CrossRef](#)]
53. Paul, S.M.; Mytelka, D.S.; Dunwiddie, C.T.; Persinger, C.C.; Munos, B.H.; Lindborg, S.R.; Schacht, A.L. How to Improve RD Productivity: The Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214. [[CrossRef](#)]
54. Ashburn, T.T.; Thor, K.B. Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683. [[CrossRef](#)] [[PubMed](#)]
55. Chong, C.R.; Sullivan, D.J. New Uses for Old Drugs. *Nature* **2007**, *448*, 645–646. [[CrossRef](#)]
56. Hernandez, J.J.; Pryszyk, M.; Smith, L.; Yanchus, C.; Kurji, N.; Shahani, V.M.; Molinski, S.V. Giving Drugs a Second Chance: Overcoming Regulatory and Financial Hurdles in Repurposing Approved Drugs As Cancer Therapeutics. *Front. Oncol.* **2017**, *7*, 273. [[CrossRef](#)] [[PubMed](#)]
57. Ma, D.L.; Chan, D.S.H.; Leung, C.H. Drug Repositioning by Structure-Based Virtual Screening. *Chem. Soc. Rev.* **2013**, *42*, 2130–2141. [[CrossRef](#)]
58. Maegawa, G.H.B.; Tropak, M.B.; Buttner, J.D.; Rigat, B.A.; Fuller, M.; Pandit, D.; Tang, L.; Kornhaber, G.J.; Hamuro, Y.; Clarke, J.T.R.; et al. Identification and Characterization of Ambroxol as an Enzyme Enhancement Agent for Gaucher Disease. *J. Biol. Chem.* **2009**, *284*, 23502–23516. [[CrossRef](#)]
59. Bendikov-Bar, I.; Maor, G.; Filocamo, M.; Horowitz, M. Ambroxol as a Pharmacological Chaperone for Mutant Glucocerebrosidase. *Blood. Cells Mol. Dis.* **2013**, *50*, 141–145. [[CrossRef](#)]
60. Narita, A.; Shirai, K.; Itamura, S.; Matsuda, A.; Ishihara, A.; Matsushita, K.; Fukuda, C.; Kubota, N.; Takayama, R.; Shigematsu, H.; et al. Ambroxol Chaperone Therapy for Neuronopathic Gaucher Disease: A Pilot Study. *Ann. Clin. Transl. Neurol.* **2016**, *3*, 200–215. [[CrossRef](#)]
61. Istiti, M.; Revel-Vilk, S.; Becker-Cohen, M.; Dinur, T.; Ramaswami, U.; Castillo-Garcia, D.; Ceron-Rodriguez, M.; Chan, A.; Rodic, P.; Tincheva, R.S.; et al. Upgrading the Evidence for the Use of Ambroxol in Gaucher Disease and GBA Related Parkinson: Investigator Initiated Registry Based on Real Life Data. *Am. J. Hematol.* **2021**, *96*, 545–551. [[CrossRef](#)]
62. Berger, A.C.; Olson, S.; Johnson, S.G.; Beachy, S.H. *Drug Repurposing and Repositioning: Workshop Summary*; Drug Repurposing and Repositioning; National Academies Press: Washington, DC, USA, 2014. [[CrossRef](#)]
63. Ng, P.C.; Henikoff, S. Predicting Deleterious Amino Acid Substitutions. *Genome Res.* **2001**, *11*, 863–874. [[CrossRef](#)] [[PubMed](#)]

64. Thomas, P.D.; Ebert, D.; Muruganujan, A.; Mushayahama, T.; Albou, L.P.; Mi, H. PANTHER: Making Genome-Scale Phylogenetics Accessible to All. *Protein Sci.* **2022**, *31*, 8–22. [[CrossRef](#)] [[PubMed](#)]
65. Snap2—Rost Lab Open. Available online: <https://roslab.org/owiki/index.php/Snap2> (accessed on 3 January 2023).
66. Calabrese, R.; Capriotti, E.; Fariselli, P.; Martelli, P.L.; Casadio, R. Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins. *Hum. Mutat.* **2009**, *30*, 1237–1244. [[CrossRef](#)] [[PubMed](#)]
67. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)]
68. Shihab, H.A.; Gough, J.; Cooper, D.N.; Stenson, P.D.; Barker, G.L.A.; Edwards, K.J.; Day, I.N.M.; Gaunt, T.R. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions Using Hidden Markov Models. *Hum. Mutat.* **2013**, *34*, 57–65. [[CrossRef](#)] [[PubMed](#)]
69. Pappalardo, M.; Wass, M.N. VarMod: Modelling the Functional Effects of Non-Synonymous Variants. *Nucleic Acids Res.* **2014**, *42*, W331–W336. [[CrossRef](#)]
70. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; et al. MutPred2: Inferring the Molecular and Phenotypic Impact of Amino Acid Variants. *Nat. Commun.* **2020**, *11*, 5918. [[CrossRef](#)] [[PubMed](#)]
71. Wong, K.C.; Zhang, Z. SNPdryad: Predicting Deleterious Non-Synonymous Human SNPs Using Only Orthologous Protein Sequences. *Bioinformatics* **2014**, *30*, 1112–1119. [[CrossRef](#)]
72. Zhou, H.; Gao, M.; Skolnick, J. ENTPRISE: An Algorithm for Predicting Human Disease-Associated Amino Acid Substitutions from Sequence Entropy and Predicted Protein Structures. *PLoS ONE* **2016**, *11*, e0150965. [[CrossRef](#)]
73. Reva, B.; Antipin, Y.; Sander, C. Determinants of Protein Function Revealed by Combinatorial Entropy Optimization. *Genome Biol.* **2007**, *8*, R232. [[CrossRef](#)]
74. Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins* **2006**, *62*, 1125–1132. [[CrossRef](#)]
75. Parthiban, V.; Gromiha, M.M.; Schomburg, D. CUPSAT: Prediction of Protein Stability upon Point Mutations. *Nucleic Acids Res.* **2006**, *34*, W239–W242. [[CrossRef](#)] [[PubMed](#)]
76. Fariselli, P.; Martelli, P.L.; Savojardo, C.; Casadio, R. INPS: Predicting the Impact of Non-Synonymous Variations on Protein Stability from Sequence. *Bioinformatics* **2015**, *31*, 2816–2821. [[CrossRef](#)] [[PubMed](#)]
77. Savojardo, C.; Fariselli, P.; Martelli, P.L.; Casadio, R. INPS-MD: A Web Server to Predict Stability of Protein Variants from Sequence and Structure. *Bioinformatics* **2016**, *32*, 2542–2544. [[CrossRef](#)]
78. Yates, C.M.; Filippis, I.; Kelley, L.A.; Sternberg, M.J.E. SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. *J. Mol. Biol.* **2014**, *426*, 2692–2701. [[CrossRef](#)]
79. Pandurangan, A.P.; Ochoa-Montaño, B.; Ascher, D.B.; Blundell, T.L. SDM: A Server for Predicting Effects of Mutations on Protein Stability. *Nucleic Acids Res.* **2017**, *45*, W229–W235. [[CrossRef](#)]
80. Rodrigues, C.H.M.; Myung, Y.; Pires, D.E.V.; Ascher, D.B. MCSM-PPI2: Predicting the Effects of Mutations on Protein–Protein Interactions. *Nucleic Acids Res.* **2019**, *47*, W338–W344. [[CrossRef](#)] [[PubMed](#)]
81. Pires, D.E.V.; Ascher, D.B.; Blundell, T.L. DUET: A Server for Predicting Effects of Mutations on Protein Stability Using an Integrated Computational Approach. *Nucleic Acids Res.* **2014**, *42*, W314–W319. [[CrossRef](#)]
82. Pires, D.E.V.; Rodrigues, C.H.M.; Ascher, D.B. MCSM-Membrane: Predicting the Effects of Mutations on Transmembrane Proteins. *Nucleic Acids Res.* **2020**, *48*, W147–W153. [[CrossRef](#)]
83. Pires, D.E.V.; Ascher, D.B. MCSM-AB: A Web Server for Predicting Antibody–Antigen Affinity Changes upon Mutation with Graph-Based Signatures. *Nucleic Acids Res.* **2016**, *44*, W469–W473. [[CrossRef](#)]
84. Rodrigues, C.H.; Pires, D.E.; Ascher, D.B.; David Ascher, C.B. DynaMut2: Assessing Changes in Stability and Flexibility upon Single and Multiple Point Missense Mutations. *Protein Sci.* **2021**, *30*, 60–69. [[CrossRef](#)] [[PubMed](#)]
85. Andreotti, G.; Guarracino, M.R.; Cammisa, M.; Correr, A.; Cubellis, M.V. Prediction of the Responsiveness to Pharmacological Chaperones: Lysosomal Human Alpha-Galactosidase, a Case of Study. *Orphanet J. Rare Dis.* **2010**, *5*, 36. [[CrossRef](#)] [[PubMed](#)]
86. Eng, C.M.; Niehaus, D.J.; Enriquez, A.L.; Burgert, T.S.; Ludman, M.D.; Desnick, R.J. Fabry Disease: Twenty-Three Mutations Including Sense and Antisense CPG Alterations and Identification of a Deletional Hot-Spot in the α -Galactosidase A Gene. *Hum. Mol. Genet.* **1994**, *3*, 1795–1799. [[CrossRef](#)]
87. du Moulin, M.; Koehn, A.F.; Golsari, A.; Dulz, S.; Atiskova, Y.; Patten, M.; Münch, J.; Avanesov, M.; Ullrich, K.; Muschol, N. The Mutation p.D313Y Is Associated with Organ Manifestation in Fabry Disease. *Clin. Genet.* **2017**, *92*, 528–533. [[CrossRef](#)]
88. Zompola, C.; Palaiodimou, L.; Kokotis, P.; Papadopoulou, M.; Theodorou, A.; Sabanis, N.; Theodoroula, E.; Papathanasiou, M.; Tsvigoulis, G.; Chroni, E. The Mutation D313Y May Be Associated with Nervous System Manifestations in Fabry Disease. *J. Neurol. Sci.* **2020**, *412*, 116757. [[CrossRef](#)]
89. Lukas, J.; Giese, A.K.; Markoff, A.; Grittner, U.; Kolodny, E.; Mascher, H.; Lackner, K.J.; Meyer, W.; Wree, P.; Saviouk, V.; et al. Functional Characterisation of Alpha-Galactosidase a Mutations as a Basis for a New Classification System in Fabry Disease. *PLoS Genet.* **2013**, *9*, e1003632. [[CrossRef](#)]
90. Koulousios, K.; Stylianou, K.; Pateinakis, P.; Zamanakou, M.; Loules, G.; Manou, E.; Kyriklidou, P.; Katsinas, C.; Ouzouni, A.; Kyriazis, J.; et al. Fabry Disease Due to D313Y and Novel GLA Mutations. *BMJ Open* **2017**, *7*, e017098. [[CrossRef](#)]

91. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [[CrossRef](#)]
92. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; Van Der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854. [[CrossRef](#)] [[PubMed](#)]
93. Case, D.A.; Cheatham, T.E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.B.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Computat. Chem.* **2005**, *26*, 1668. [[CrossRef](#)]
94. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hémin, J.; Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [[CrossRef](#)] [[PubMed](#)]
95. Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossváry, I.; Moraes, M.; Sacerdoti, F.; et al. Molecular Dynamics—Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In Proceedings of the SC'06, 2006 ACM/IEEE Conference on Supercomputing, Tampa, FL, USA, 11–17 November 2006; p. 84.
96. Rackers, J.A.; Wang, Z.; Lu, C.; Laury, M.L.; Lagardère, L.; Schnieders, M.J.; Piquemal, J.P.; Ren, P.; Ponder, J.W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory. Comput.* **2018**, *14*, 5273–5289. [[CrossRef](#)] [[PubMed](#)]
97. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
98. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version~1.8*; 2015.
99. Boyce, S.E.; Mobley, D.L.; Rocklin, G.J.; Graves, A.P.; Dill, K.A.; Shoichet, B.K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009**, *394*, 747–763. [[CrossRef](#)]
100. Lundborg, M.; Lidmar, J.; Hess, B. The Accelerated Weight Histogram Method for Alchemical Free Energy Calculations. *J. Chem. Phys.* **2021**, *154*, 204103. [[CrossRef](#)] [[PubMed](#)]
101. Bennett, C.H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268. [[CrossRef](#)]
102. Matsunaga, Y.; Kamiya, M.; Oshima, H.; Jung, J.; Ito, S.; Sugita, Y. Use of Multistate Bennett Acceptance Ratio Method for Free-Energy Calculations from Enhanced Sampling and Free-Energy Perturbation. *Biophys. Rev.* **2022**, *14*, 1503–1512. [[CrossRef](#)]
103. Lindahl, V.; Lidmar, J.; Hess, B. Accelerated Weight Histogram Method for Exploring Free Energy Landscapes. *J. Chem. Phys.* **2014**, *141*, 044110. [[CrossRef](#)]
104. Daggett, V. Molecular Dynamics Simulations of the Protein Unfolding/Folding Reaction. *Acc. Chem. Res.* **2002**, *35*, 422–429. [[CrossRef](#)]
105. Childers, M.C.; Daggett, V. Validating Molecular Dynamics Simulations against Experimental Observables in Light of Underlying Conformational Ensembles. *J. Phys. Chem. B* **2018**, *122*, 6673–6689. [[CrossRef](#)]
106. Mayor, U.; Johnson, C.M.; Daggett, V.; Fersht, A.R.; Kazmirski, S.L.; Wong, K.B.; Freund, S.M.V.; Tan, Y.-J. Protein Folding and Unfolding in Microseconds to Nanoseconds by Experiment and Simulation. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13518–13522. [[CrossRef](#)] [[PubMed](#)]
107. Zhang, D.; Lazim, R. Application of Conventional Molecular Dynamics Simulation in Evaluating the Stability of Apomyoglobin in Urea Solution. *Sci. Rep.* **2017**, *7*, 44651. [[CrossRef](#)]
108. Dechene, M.; Wink, G.; Smith, M.; Swartz, P.; Mattos, C. Multiple Solvent Crystal Structures of Ribonuclease A: An Assessment of the Method. *Proteins Struct. Funct. Bioinform.* **2009**, *76*, 861–881. [[CrossRef](#)] [[PubMed](#)]
109. Agarwal, S.; Smith, M.; De La Rosa, I.; Verba, K.A.; Swartz, P.; Segura-Totten, M.; Mattos, C. Development of a Structure-Analysis Pipeline Using Multiple-Solvent Crystal Structures of Barrier-to-Autointegration Factor. *Acta Cryst. D Struct. Biol.* **2020**, *76*, 1001–1014. [[CrossRef](#)]
110. Mattos, C.; Bellamacina, C.R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G.A.; Ringe, D. Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration. *J. Mol. Biol.* **2006**, *357*, 1471–1482. [[CrossRef](#)]
111. Nguyen, T.T.; Venditti, V. An Allosteric Pocket for Inhibition of Bacterial Enzyme I Identified by NMR-Based Fragment Screening. *J. Struct. Biol. X* **2020**, *4*, 100034. [[CrossRef](#)]
112. Bernini, A.; Venditti, V.; Spiga, O.; Niccolai, N. Probing Protein Surface Accessibility with Solvent and Paramagnetic Molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *54*, 278–289. [[CrossRef](#)]
113. Consortium, T.U.; Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; Bye-A-Jee, H.; et al. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2013**, *1*, 13–14. [[CrossRef](#)]
114. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
115. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, *46*, W363–W367. [[CrossRef](#)] [[PubMed](#)]
116. McGreig, J.E.; Uri, H.; Antczak, M.; Sternberg, M.J.E.; Michaelis, M.; Wass, M.N. 3DLigandSite: Structure-Based Prediction of Protein–Ligand Binding Sites. *Nucleic Acids Res.* **2022**, *50*, W13–W20. [[CrossRef](#)]
117. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]

118. McGuffin, L.J.; Adiyaman, R.; Maghrabi, A.H.A.; Shuid, A.N.; Brackenridge, D.A.; Nealon, J.O.; Philomina, L.S. IntFOLD: An Integrated Web Resource for High Performance Protein Structure and Function Prediction. *Nucleic Acids Res.* **2019**, *47*, W408–W413. [[CrossRef](#)] [[PubMed](#)]
119. Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A.S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33*, 3036–3042. [[CrossRef](#)] [[PubMed](#)]
120. Wu, Q.; Peng, Z.; Zhang, Y.; Yang, J. COACH-D: Improved Protein-Ligand Binding Sites Prediction with Refined Ligand-Binding Poses through Molecular Docking. *Nucleic Acids Res.* **2018**, *46*, W438–W442. [[CrossRef](#)] [[PubMed](#)]
121. Jendele, L.; Krivak, R.; Skoda, P.; Novotny, M.; Hoksza, D. PrankWeb: A Web Server for Ligand Binding Site Prediction and Visualization. *Nucleic Acids Res.* **2019**, *47*, W345–W349. [[CrossRef](#)] [[PubMed](#)]
122. Citro, V.; Peña-García, J.; Den-Haan, H.; Pérez-Sánchez, H.; Del Prete, R.; Liguori, L.; Cimmaruta, C.; Lukas, J.; Cubellis, M.V.; Andreotti, G. Identification of an Allosteric Binding Site on Human Lysosomal Alpha-Galactosidase Opens the Way to New Pharmacological Chaperones for Fabry Disease. *PLoS ONE* **2016**, *11*, e0165463. [[CrossRef](#)] [[PubMed](#)]
123. Benz, J.; Rufer, A.C.; Huber, S.; Ehler, A.; Hug, M.; Topp, A.; Guba, W.; Hofmann, E.C.; Jagasia, R.; Rodríguez Sarmiento, R.M. Novel β -Glucocerebrosidase Activators That Bind to a New Pocket at a Dimer Interface and Induce Dimerization. *Angew. Chem. Int. Ed.* **2021**, *60*, 5436–5442. [[CrossRef](#)]
124. Oleinikovas, V.; Saladino, G.; Cossins, B.P.; Gervasio, F.L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 14257–14263. [[CrossRef](#)] [[PubMed](#)]
125. Wells, J.A.; McClendon, C.L. Reaching for High-Hanging Fruit in Drug Discovery at Protein–Protein Interfaces. *Nature* **2007**, *450*, 1001–1009. [[CrossRef](#)] [[PubMed](#)]
126. Eyrich, S.; Helms, V. Transient Pockets on Protein Surfaces Involved in Protein-Protein Interaction. *J. Med. Chem.* **2007**, *50*, 3457–3464. [[CrossRef](#)]
127. Seddon, G.M.; Bywater, R.P. Accelerated Simulation of Unfolding and Refolding of a Large Single Chain Globular Protein. *Open Biol.* **2012**, *2*, 120087. [[CrossRef](#)]
128. Eyrich, S.; Helms, V. What Induces Pocket Openings on Protein Surface Patches Involved in Protein-Protein Interactions? *J. Comput. Aided Mol. Des.* **2009**, *23*, 73–86. [[CrossRef](#)]
129. Joerger, A.C.; Bauer, M.R.; Wilcken, R.; Baud, M.G.J.; Harbrecht, H.; Exner, T.E.; Boeckler, F.M.; Spencer, J.; Fersht, A.R. Exploiting Transient Protein States for the Design of Small-Molecule Stabilizers of Mutant P53. *Structure* **2015**, *23*, 2246–2255. [[CrossRef](#)]
130. Wassman, C.D.; Baronio, R.; Demir, Ö.; Wallentine, B.D.; Chen, C.K.; Hall, L.V.; Salehi, F.; Lin, D.W.; Chung, B.P.; Wesley Hatfield, G.; et al. Computational Identification of a Transiently Open L1/S3 Pocket for Reactivation of Mutant P53. *Nat. Commun.* **2013**, *4*, 1407. [[CrossRef](#)] [[PubMed](#)]
131. Bernini, A.; Henrici De Angelis, L.; Morandi, E.; Spiga, O.; Santucci, A.; Assalg, M.; Molinari, H.; Pillozzi, S.; Arcangeli, A.; Niccolai, N. Searching for Protein Binding Sites from Molecular Dynamics Simulations and Paramagnetic Fragment-Based NMR Studies. *Biochim. Biophys. Acta* **2014**, *1844*, 561–566. [[CrossRef](#)] [[PubMed](#)]
132. Pietrucci, F.; Vargiu, A.V.; Kranjc, A. HIV-1 Protease Dimerization Dynamics Reveals a Transient Druggable Binding Pocket at the Interface. *Sci. Rep.* **2015**, *5*, 18555. [[CrossRef](#)]
133. Bernini, A.; Galderisi, S.; Spiga, O.; Bernardini, G.; Niccolai, N.; Manetti, F.; Santucci, A. Toward a Generalized Computational Workflow for Exploiting Transient Pockets as New Targets for Small Molecule Stabilizers: Application to the Homogentisate 1,2-Dioxygenase Mutants at the Base of Rare Disease Alkaptonuria. *Comput. Biol. Chem.* **2017**, *70*, 133–141. [[CrossRef](#)] [[PubMed](#)]
134. Schmidtke, P.; Bidon-chanal, A.; Luque, F.J.; Barril, X. MDpocket: Open-Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27*, 3276–3285. [[CrossRef](#)]
135. Brady, G.P.; Stouten, P.F.W. Fast Prediction and Visualization of Protein Binding Pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14*, 383–401. [[CrossRef](#)] [[PubMed](#)]
136. Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tufféry, P. Fpocket: Online Tools for Protein Ensemble Pocket Detection and Tracking. *Nucleic Acids Res.* **2010**, *38*, W582. [[CrossRef](#)] [[PubMed](#)]
137. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [[CrossRef](#)] [[PubMed](#)]
138. Shultz, M.D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62*, 1701–1714. [[CrossRef](#)]
139. Hussein, H.A.; Borrel, A.; Geneix, C.; Petitjean, M.; Regad, L.; Camproux, A.C. PockDrug-Server: A New Web Server for Predicting Pocket Druggability on Holo and Apo Proteins. *Nucleic Acids Res.* **2015**, *43*, W436–W442. [[CrossRef](#)] [[PubMed](#)]
140. Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R. DrugPred: A Structure-Based Approach to Predict Protein Druggability Developed Using an Extensive Nonredundant Data Set. *J. Chem. Inf. Model.* **2011**, *51*, 2829–2842. [[CrossRef](#)] [[PubMed](#)]
141. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A Web Server for Automatic Binding Site Prediction, Analysis and Druggability Assessment. *Bioinformatics* **2012**, *28*, 2074–2075. [[CrossRef](#)] [[PubMed](#)]
142. Anand, P.; Nagarajan, D.; Mukherjee, S.; Chandra, N. PLIC: Protein–Ligand Interaction Clusters. *Database* **2014**, *2014*, bau029. [[CrossRef](#)]
143. Van Dongen, S.; Abreu-Goodger, C. Using MCL to Extract Clusters from Networks. *Methods Mol. Biol.* **2012**, *804*, 281–295.

144. Sobolev, V.; Eyal, E.; Gerzon, S.; Potapov, V.; Babor, M.; Prilusky, J.; Edelman, M. SPACE: A Suite of Tools for Protein Structure Prediction and Analysis Based on Complementarity and Environment. *Nucleic Acids Res.* **2005**, *33*, W39–W43. [[CrossRef](#)]
145. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comp. Chem.* **1998**, *19*, 1639–1662. [[CrossRef](#)]
146. Xu, Y.; Wang, S.; Hu, Q.; Gao, S.; Ma, X.; Zhang, W.; Shen, Y.; Chen, F.; Lai, L.; Pei, J. CavityPlus: A Web Server for Protein Cavity Detection with Pharmacophore Modelling, Allosteric Site Identification and Covalent Ligand Binding Ability Prediction. *Nucleic Acids Res.* **2018**, *46*, W374–W379. [[CrossRef](#)] [[PubMed](#)]
147. Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H. PharmMapper 2017 Update: A Web Server for Potential Drug Target Identification with a Comprehensive Target Pharmacophore Database. *Nucleic Acids Res.* **2017**, *45*, W356–W360. [[CrossRef](#)] [[PubMed](#)]
148. Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper Server: A Web Server for Potential Drug Target Identification Using Pharmacophore Mapping Approach. *Nucleic Acids Res.* **2010**, *38*, W609. [[CrossRef](#)]
149. Bender, B.J.; Gahbauer, S.; Lutten, A.; Lyu, J.; Webb, C.M.; Stein, R.M.; Fink, E.A.; Balias, T.E.; Carlsson, J.; Irwin, J.J.; et al. A Practical Guide to Large-Scale Docking. *Nat. Protoc.* **2021**, *16*, 4799–4832. [[CrossRef](#)] [[PubMed](#)]
150. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J. Med. Chem.* **2013**, *56*, 6560–6572. [[CrossRef](#)]
151. Goodwin, S.; Shahtahmassebi, G.; Hanley, Q.S. Statistical Models for Identifying Frequent Hitters in High Throughput Screening. *Sci. Rep.* **2020**, *10*, 17200. [[CrossRef](#)] [[PubMed](#)]
152. Irwin, J.J.; Tang, K.G.; Young, J.; Dandarchuluun, C.; Wong, B.R.; Khurelbaatar, M.; Moroz, Y.S.; Mayfield, J.; Sayle, R.A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073. [[CrossRef](#)]
153. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [[CrossRef](#)]
154. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
155. Gaulton, A.; Hersey, A.; Nowotka, M.L.; Patricia Bento, A.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; et al. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [[CrossRef](#)]
156. Pihan, E.; Colliandre, L.; Guichou, J.F.; Douguet, D. E-Drug3D: 3D Structure Collections Dedicated to Drug Repurposing and Fragment-Based Drug Design. *Bioinformatics* **2012**, *28*, 1540–1541. [[CrossRef](#)] [[PubMed](#)]
157. Siramshetty, V.B.; Eckert, O.A.; Gohlke, B.O.; Goede, A.; Chen, Q.; Devarakonda, P.; Preissner, S.; Preissner, R. SuperDRUG2: A One Stop Resource for Approved/Marketed Drugs. *Nucleic Acids Res.* **2018**, *46*, D1137–D1143. [[CrossRef](#)]
158. Chen, X.; Liu, M.; Gilson, M. BindingDB: A Web-Accessible Molecular Recognition Database. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 719–725. [[CrossRef](#)]
159. Wishart, D.S.; Guo, A.C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [[CrossRef](#)]
160. Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, M. LIGAND: Database of Chemical Compounds and Reactions in Biological Pathways. *Nucleic Acids Res.* **2002**, *30*, 402–404. [[CrossRef](#)] [[PubMed](#)]
161. Grygorenko, O.O.; Radchenko, D.S.; Dziuba, I.; Chuprina, A.; Gubina, K.E.; Moroz, Y.S. Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. *iScience* **2020**, *23*, 101681. [[CrossRef](#)]
162. Ruddigkeit, L.; Van Deursen, R.; Blum, L.C.; Raymond, J.L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875. [[CrossRef](#)] [[PubMed](#)]
163. Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [[CrossRef](#)]
164. Santos, L.H.S.; Ferreira, R.S.; Caffarena, E.R. Integrating Molecular Docking and Molecular Dynamics Simulations. *Methods Mol. Biol.* **2019**, *2053*, 13–34.
165. Cerón-Carrasco, J.P.; Mozzarelli, A.; Moro, S.; Salmaso, V. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.* **1980**, *1*, 923. [[CrossRef](#)]
166. Morris, G.M.; Ruth, H.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785. [[CrossRef](#)] [[PubMed](#)]
167. Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152. [[CrossRef](#)]
168. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [[CrossRef](#)] [[PubMed](#)]
169. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
170. Ravindranath, P.A.; Forli, S.; Goodsell, D.S.; Olson, A.J.; Sanner, M.F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput. Biol.* **2015**, *11*, e1004586. [[CrossRef](#)] [[PubMed](#)]
171. McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578–596. [[CrossRef](#)]

172. Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A.R.; Ansbacher, T.; Major, D.T. EnzyDock: Protein-Ligand Docking of Multiple Reactive States along a Reaction Coordinate in Enzymes. *J. Chem. Theory Comput.* **2019**, *15*, 5116–5134. [CrossRef] [PubMed]
173. Flexible Protein–Ligand Docking by Global Energy Optimization in Internal Coordinates—Totrov-1997-Proteins: Structure, Function, and Bioinformatics—Wiley Online Library. Available online: [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0134\(1997\)1+%3C215::AID-PROT29%3E3.0.CO;2-Q](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0134(1997)1+%3C215::AID-PROT29%3E3.0.CO;2-Q) (accessed on 31 January 2023).
174. Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX Incremental Construction Algorithm for Protein-Ligand Docking. *Proteins Struct. Funct. Bioinform.* **1999**, *37*, 228–241. [CrossRef]
175. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [CrossRef]
176. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved Protein–Ligand Docking Using GOLD. *Proteins Struct. Funct. Bioinform.* **2003**, *52*, 609–623. [CrossRef] [PubMed]
177. Huang, S.Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins Struct. Funct. Bioinform.* **2007**, *66*, 399–421. [CrossRef]
178. Vilar, S.; Cozza, G.; Moro, S. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Curr. Top. Med. Chem.* **2008**, *8*, 1555–1572. [CrossRef]
179. Murrugan, N.A.; Podobas, A.; Gadioli, D.; Vitali, E.; Palermo, G.; Markidis, S. A Review on Parallel Virtual Screening Softwares for High-Performance Computers. *Pharmaceuticals* **2022**, *15*, 63. [CrossRef]
180. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9*, 91. [CrossRef]
181. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. 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*, 12964–12975. [CrossRef] [PubMed]
182. Cross, J.B.; Thompson, D.C.; Rai, B.K.; Baber, J.C.; Fan, K.Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455–1474. [CrossRef]
183. Çınaroğlu, S.S.; Timuçin, E. Comparative Assessment of Seven Docking Programs on a Nonredundant Metalloprotein Subset of the PDBbind Refined. *J. Chem. Inf. Model.* **2019**, *59*, 3846–3859. [CrossRef] [PubMed]
184. Forli, S.; Huey, R.; Pique, M.E.; Sanner, M.F.; Goodsell, D.S.; Olson, A.J. Computational Protein–Ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11*, 905–919. [CrossRef] [PubMed]
185. Chen, Y.C. Beware of Docking! *Trends. Pharm. Sci.* **2015**, *36*, 78–95. [CrossRef]
186. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J.Z.H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478–9508. [CrossRef]
187. Thirumal Kumar, D.; Iyer, S.; Christy, J.P.; Siva, R.; Tayubi, I.A.; George Priya Doss, C.; Zayed, H. A Comparative Computational Approach toward Pharmacological Chaperones (NN-DNJ and Ambroxol) on N370S and L444P Mutations Causing Gaucher’s Disease. *Adv. Protein Chem. Struct. Biol.* **2019**, *114*, 315–339. [CrossRef] [PubMed]
188. Nakagome, I.; Kato, A.; Yamaotsu, N.; Yoshida, T.; Ozawa, S.I.; Adachi, I.; Hirono, S.S. Design of a New α -1-C-Alkyl-DAB Derivative Acting as a Pharmacological Chaperone for β -Glucocerebrosidase Using Ligand Docking and Molecular Dynamics Simulation. *Molecules* **2018**, *23*, 2683. [CrossRef] [PubMed]
189. Yilmazer, B.; Yagci, Z.B.; Bakar, E.; Ozden, B.; Ulgen, K.; Ozkirimli, E. Investigation of Novel Pharmacological Chaperones for Gaucher Disease. *J. Mol. Graph. Model.* **2017**, *76*, 364–378. [CrossRef]
190. Kato, A.; Nakagome, I.; Sato, K.; Yamamoto, A.; Adachi, I.; Nash, R.J.; Fleet, G.W.J.; Natori, Y.; Watanabe, Y.; Imahori, T.; et al. Docking Study and Biological Evaluation of Pyrrolidine-Based Iminosugars as Pharmacological Chaperones for Gaucher Disease. *Org. Biomol. Chem.* **2016**, *14*, 1039–1048. [CrossRef]
191. Kato, A.; Nakagome, I.; Nakagawa, S.; Koike, Y.; Nash, R.J.; Adachi, I.; Hirono, S. Docking and SAR Studies of Calystegines: Binding Orientation and Influence on Pharmacological Chaperone Effects for Gaucher’s Disease. *Bioorg. Med. Chem.* **2014**, *22*, 2435–2441. [CrossRef]
192. Kato, A.; Nakagome, I.; Nakagawa, S.; Kinami, K.; Adachi, I.; Jenkinson, S.F.; Désiré, J.; Blériot, Y.; Nash, R.J.; Fleet, G.W.J.; et al. In Silico Analyses of Essential Interactions of Iminosugars with the Hex A Active Site and Evaluation of Their Pharmacological Chaperone Effects for Tay-Sachs Disease. *Org. Biomol. Chem.* **2017**, *15*, 9297–9304. [CrossRef] [PubMed]
193. Olarte-Avellaneda, S.; Cepeda Del Castillo, J.; Rojas-Rodríguez, A.F.; Sánchez, O.; Rodríguez-López, A.; Suárez García, D.A.; Pulido, L.M.S.; Alméciga-Díaz, C.J. Bromocriptine as a Novel Pharmacological Chaperone for Mucopolysaccharidosis IV A. *ACS Med. Chem. Lett.* **2020**, *11*, 1377–1385. [CrossRef] [PubMed]
194. Ballester, P.J.; Mitchell, J.B.O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169. [CrossRef] [PubMed]
195. Yang, C.; Chen, E.A.; Zhang, Y. Protein–Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, *27*, 4568. [CrossRef]
196. Crampon, K.; Giorkallos, A.; Deldossi, M.; Baud, S.; Steffanel, L.A. Machine-Learning Methods for Ligand–Protein Molecular Docking. *Drug Discov. Today* **2022**, *27*, 151–164. [CrossRef]
197. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva-Jr, F.P. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [CrossRef]

198. Amendola, G.; Cosconati, S. PyRMD: A New Fully Automated AI-Powered Ligand-Based Virtual Screening Tool. *J. Chem. Inf. Model.* **2021**, *61*, 3835–3845. [[CrossRef](#)]
199. Kumar, N.; Acharya, V. Machine Intelligence-Driven Framework for Optimized Hit Selection in Virtual Screening. *J. Cheminform.* **2022**, *14*, 48. [[CrossRef](#)]
200. Tripathi, M.K.; Nath, A.; Singh, T.P.; Ethayathulla, A.S.; Kaur, P. Evolving Scenario of Big Data and Artificial Intelligence (AI) in Drug Discovery. *Mol. Divers.* **2021**, *25*, 1439–1460. [[CrossRef](#)] [[PubMed](#)]
201. Arul Murugan, N.; Ruba Priya, G.; Narahari Sastry, G.; Markidis, S. Artificial Intelligence in Virtual Screening: Models versus Experiments. *Drug Discov. Today* **2022**, *27*, 1913–1923. [[CrossRef](#)]
202. Böhm, H.J.; Flohr, A.; Stahl, M. Scaffold Hopping. *Drug Discov. Today Technol.* **2004**, *1*, 217–224. [[CrossRef](#)]
203. Bajorath, J. Computational Scaffold Hopping: Cornerstone for the Future of Drug Design? *Future Med. Chem.* **2017**, *9*, 629–631. [[CrossRef](#)] [[PubMed](#)]
204. Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.T.; Ban, F.; Norinder, U.; Gleave, M.E.; Cherkasov, A. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS. Cent. Sci.* **2020**, *6*, 939–949. [[CrossRef](#)]
205. Yaacoub, J.C.; Gleave, J.; Gentile, F.; Stern, A.; Cherkasov, A. DD-GUI: A Graphical User Interface for Deep Learning-Accelerated Virtual Screening of Large Chemical Libraries (Deep Docking). *Bioinformatics* **2022**, *38*, 1146–1148. [[CrossRef](#)]
206. Graff, D.E.; Shakhnovich, E.I.; Coley, C.W. Accelerating High-Throughput Virtual Screening through Molecular Pool-Based Active Learning. *Chem. Sci.* **2021**, *12*, 7866–7881. [[CrossRef](#)] [[PubMed](#)]
207. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.F.; Fischer, P.D.; Coote, P.W.; Padmanabha Das, K.M.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A.; et al. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature* **2020**, *580*, 663–668. [[CrossRef](#)]
208. Gorgulla, C.; Çinaroğlu, S.S.; Fischer, P.D.; Fackeldey, K.; Wagner, G.; Arthanari, H. VirtualFlow Ants—Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization. *Int. J. Mol. Sci.* **2021**, *22*, 5807. [[CrossRef](#)] [[PubMed](#)]
209. Sadybekov, A.A.; Sadybekov, A.V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N.K.; Tong, F.; et al. Synthon-Based Ligand Discovery in Virtual Libraries of over 11 Billion Compounds. *Nature* **2022**, *601*, 452–459. [[CrossRef](#)]
210. Schneider, G.; Clark, D.E. Automated De Novo Drug Design: Are We Nearly There Yet? *Angew. Chem. Int. Ed.* **2019**, *58*, 10792–10803. [[CrossRef](#)] [[PubMed](#)]
211. Merk, D.; Friedrich, L.; Grisoni, F.; Schneider, G. De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Mol. Inform.* **2018**, *37*, 1700153. [[CrossRef](#)]
212. Lee, Y.J.; Kahng, H.; Kim, S.B. Generative Adversarial Networks for De Novo Molecular Design. *Mol. Inform.* **2021**, *40*, 2100045. [[CrossRef](#)] [[PubMed](#)]
213. Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* **2018**, *4*, eaap7885. [[CrossRef](#)]
214. Palazzesi, F.; Pozzan, A. Deep Learning Applied to Ligand-Based De Novo Drug Design. *Methods Mol. Biol.* **2022**, *2390*, 273–299. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.