*Review*

# Maximum Entropy in Drug Discovery

**Chih-Yuan Tseng [1,†,\*] and Jack Tuszynski [1,2]**

[1] Department of Oncology, University of Alberta, Edmonton, AB T6G 1Z2, Canada
[2] Department of Physics, University of Alberta, Edmonton, AB T6G 1Z2, Canada;
E-Mail: jackt@ualberta.ca
[†] Current address: Sinoveda Canada Inc, Edmonton, AB T6N 1H1, Canada

[\*] Author to whom correspondence should be addressed; E-Mail: chih-yuan.tseng@ualberta.ca;
Tel.:+1-780-9082104.

**Abstract:** Drug discovery applies multidisciplinary approaches either experimentally, computationally or both ways to identify lead compounds to treat various diseases. While conventional approaches have yielded many US Food and Drug Administration (FDA)-approved drugs, researchers continue investigating and designing better approaches to increase the success rate in the discovery process. In this article, we provide an overview of the current strategies and point out where and how the method of maximum entropy has been introduced in this area. The maximum entropy principle has its root in thermodynamics, yet since Jaynes' pioneering work in the 1950s, the maximum entropy principle has not only been used as a physics law, but also as a reasoning tool that allows us to process information in hand with the least bias. Its applicability in various disciplines has been abundantly demonstrated. We give several examples of applications of maximum entropy in different stages of drug discovery. Finally, we discuss a promising new direction in drug discovery that is likely to hinge on the ways of utilizing maximum entropy.

**Keywords:** maximum entropy; inductive inference; drug discovery; target identification; compound design; pharmacokinetics; pharmacodynamics

**PACS Codes:** 02.50.Tt; 87.10.Vg

## 1. Introduction

Drug discovery applies multidisciplinary approaches to identify lead compounds to treat various diseases either experimentally, computationally or using both types of approaches. In this context, one faces an issue that information required to "discover" drugs is always insufficient because of the inherent complexity of biological systems. Consequently, researchers continue investigating and designing better approaches to improve the success rate in the discovery process. Since the method of maximum entropy is designed to solve problems for the cases that have insufficient information, we argue that maximum entropy may provide a very appropriate approach to drug discovery. However, as shown later, it is surprising that we only can discover a limited number of applications of entropy-based methods to various topics in drug discovery. The reason may be because the development of the maximum entropy method as a tool for inductive inference was roughly conducted over a similar time span when computational rational drug discovery was developed. Namely, researchers paid more attention to the foundation of the maximum entropy method and its applications in fields such as astronomy, physics and engineering. Nevertheless, we believe these examples provide sufficient support to demonstrate the use of maximum entropy.

In this review article, we will not only briefly discuss how the maximum entropy principle is applied to solving problems emerging in drug discovery but also how it is an attempt to provide an alternative but more general strategy compared to standard methods. This strategy is what we later refer to as the entropy-based scheme. We hope the illustration here reveals advantages of this scheme and the way of utilizing it to tackle different topics in drug discovery.

This article is arranged in the following order: in the next section, we discuss the maximum entropy method and define the entropy-based scheme. In Section 3, we discuss the kinds of topics and problems one faces in drug discovery. In Section 4, for each topic discussed in Section 3, we show how entropy-based methods can be used to solve problems that are typically encountered in these situations. Furthermore, we show a common reasoning framework that is not obvious in the original work behind all of these applications. Namely, it is the entropy-based scheme. Finally, we provide conclusions and future outlook.

## 2. Maximum Entropy Principle and Inductive Inference

Two reasoning methods have been utilized in developing theories to interpret phenomena we observe in Nature and to make predictions about them. The first is deduction. It allows us to draw conclusions when sufficient information is available. The second method is variously called inductive logic, inductive inference, probable inference or just inference. It allows us to reason when the available information is insufficient for deduction. Namely, it is called either "inference" for the cases that we make estimates of quantities when we do not know enough to deduce their values, or "induction" when we generalize from special cases [1].

When dealing with complicated biological systems, which involve either many-body interactions at a microscopic level, complicated regulating protein-protein networks at a mesoscopic level or genetic populations at a macroscopic level, it is common to find that we do not have sufficient knowledge to adequately understand these kinds of systems. We normally rely on inductive inference to deduce the

most preferable solutions to the problems related to these systems based on available information. As will be discussed later, the field of drug discovery is one of such examples.

In this section, we discuss the mathematical tools required for inductive inference. The idea hinges on the Bayesian interpretation of probability and the rules of probability theory. The former treats probability as a measure of our knowledge about the system of interest rather than a frequency of occurrence of an event. The latter demonstrates that this type of probability can be manipulated by the rules of subtraction, multiplication and addition given by the standard probability theory [2,3]. These two tools form the building blocks for inductive inference and its core is the concept of entropy.

*The state of knowledge.* Based on Laplace's principle of insufficient reasoning, Bayesian interpretation of probability and Cox's axiomatic approach, Jaynes first showed that the method of maximum entropy (denoted here by MaxEnt for short) is a reasoning tool to assign probabilities on the basis of limited information about the systems of interest with the least bias [4,5]. The preferred probability distribution of the system in a specific state is the one that maximizes the entropy of the system subject to the constraint of a limited amount of information available. Note that entropy is defined by the well-known formula:

$$S[p] = -\sum_i p_i \ln p_i \tag{1}$$

where $p_i$ is the probability of the system in state $i$ and the Boltzmann constant has been set to one for convenience. And the "information" about the system available is in the form of constraints as described by:

$$\langle O \rangle = -\sum_i p_i O_i \ , \tag{2}$$

which represents the expectation value of the observed quantity $O_i$. The probability distribution $p_i$ represents our state of knowledge regarding the system in state $i$. In summary, Jaynes' pioneering work established a mathematical framework for determining the most preferable inference, namely the probability one can estimate according to the information about the system in hand. Entropy is not just a physical quantity for measuring the degree of randomness or information. It is a reasoning tool, too.

*Maximum entropy based inference scheme.* The next breakthrough was to establish a tool beyond MaxEnt for updating our state of knowledge regarding the system based on new information acquired. Shore and Johnson, Skilling and Caticha followed the same reasoning as Jaynes did and proposed that the method of maximum entropy (denoted here by ME) can also be utilized as a tool for updating probability distributions based on new information [4,6–9]. They showed that the key to updating probability distributions is through relative entropy rather than the entropy used in MaxEnt:

$$S[p,m] = -\sum_i p_i \ln {p_i}/{m_i} \tag{3}$$

where $m_i$ an arbitrary prior, is a probability distribution regarding the system in state $i$ before any new information is acquired. Caticha further demonstrated that ME unifies the two methods, MaxEnt and Bayesian methods into a single inference scheme [10]. In summary, based on this scheme, inductive inference is made by first resolving the following two issues, first: what is a prior and hence what information is relevant to the system of interest, and second: what is the relevant application of the

method of maximum entropy. Since ME allows arbitrary prior and constraints, one can expect this general scheme can be applied to any kinds of problems as long as information acquired to investigate these problems is in the form of constraints.

*General inference scheme*. As discussed above, the most probable inference made by ME is the one that maximizes relative entropy, Equation (3). However, the question remains to what extent are less favorable inferences ruled out by the maximum entropy principle. Information geometry (IG) offers a straight-forward tool to answer this question [11]. IG introduces the concept of manifolds for characterizing and manipulating information [12,13]. Probability distributions then can be simply treated as points on the information manifold, which is defined based on the Fisher-Rao metric [12]. Therefore, one can quantify preferences for all probability distributions, inferences, through a distribution of corresponding points on the information manifold [11]. Utilizing ME and IG together then provides a general scheme to manipulate information and make inductive inference [11]. Namely, one will not only make the most favorable inference but also has the knowledge regarding the preference of less favorable inferences.

*Comment.* In summary, the entropy-based scheme is a reasoning framework based on the principle of maximum entropy. MaxEnt, ME and IG provide mathematical tools to manipulate a belief about the system under investigation utilizing a probability distribution. The most important thing in applying this scheme to solve problems is to ask "the right question". We will illustrate this statement in more detail when discussing various examples relevant to drug discovery.

## 3. Rational Drug Discovery

Before we explore the use of the entropy-based scheme in drug discovery, we discuss how a drug is typically "discovered". The central idea of the function of a drug is to inhibit (or sometimes activate) functions of biological targets (proteins, enzymes or DNA) in our body to prevent or cure diseases that are related to abnormal activities of such targets. The conventional wisdom employed in an effort to "discover" such a drug includes the following five aspects: (a) identification of biological targets; (b) design of small molecules to bind to active sites of targets; (c) screening of these molecules to determine candidates that satisfy specific properties to act as a drug; (d) demonstration of the desired pharmacokinetics of these agents and (e) demonstration of sufficient potency. We briefly discuss each subject in the following two sections.

*3.1. Compound Design*

3.1.1. Target Identification

The first stage in drug discovery is to identify biological targets that are responsible for causing abnormal physiological activities in biological systems. In the past, the complexity of biological systems has limited the concept of drug design to the so-called single drug-single target paradigm. However, there is increasing evidence that this paradigm fails to yield better drugs with satisfactory effects expected to treat diseases [14–17]. The recent blossoming of the field of systems biology further demonstrates that this paradigm is inadequate because of the enormous complexity of

signaling/regulatory networks involved in the emergence and stability of disease states [18,19]. Namely, there may be multiple targets in the networks involved in cellular state regulation. These studies thus further suggest a better paradigm whose central idea is rooted in the multiple-target multiple-drug principle [18,19]. Systems biology based approaches therefore were devised to assess which targets are likely to be key factors to inhibit abnormal regulations caused by diseases. The method proposed by Yang *et al.* [19] is one specific example illustrating this point. These authors propose to perturb the network and optimize it toward the desired state using a Monte Carlo simulated annealing optimization algorithm. By comparing diseased and normal network states, one can assess which targets are most likely responsible for the disease states as potential drug targets.

### 3.1.2. Determination of Target Structures

After key biological targets are identified, one needs to determine their tertiary structures in order to design small molecules that can specifically and selectively bind to active sites on them to regulate their functions. Several wet lab techniques including NMR and X-ray crystallography have been the primary tools for this purpose. If target structures cannot be determined through either NMR or X-ray techniques, one still can apply the homology modeling method to construct a reliable approximation to the tertiary structure of the target molecule [20]. The homology modeling method involves applications of several bioinformatic tools including protein sequence analysis and structure analysis (it consists of comparison, construction, equilibration and validation). The reader is referred to Reference [20] for more details.

### 3.1.3. Design Strategy

The size and structural diversity of the library of chemical compounds used in the process are key factors to successfully discover lead compounds that can bind specifically to molecular targets. However, the chemical space covering organic molecules is astronomically large. It is highly unlikely that one can efficiently discover lead compounds through screening compound libraries without introducing any efficient searching algorithms. In the past three decades, virtual screening approaches based on integration of various computational methods have been developed to overcome this issue [21,22]. Basically, these approaches consist of three ingredients, namely the creation of a chemical library; search for an optimal ligand-receptor binding mode through docking algorithms; and evaluation of the binding preference. Accordingly, three criteria are required to successfully identify compound candidates. First, the chemical library needs to contain large numbers of diverse chemical structures. Second, conformational searching algorithms need to be able to screen all possible binding modes within reasonable time. Third, an appropriate scoring function needs to be utilized to correctly evaluate the binding preference of specific compounds. Autodock software is one of the methods commonly used in virtual screening approaches. It utilizes a grid-space method for rapid binding energy calculation, a Lamarckian genetic algorithm for increasing efficiency and accuracy in conformational searches and a semi-empirical free energy force field to more precisely evaluate the binding free energy [23]. With the advance of screening techniques and energy calculation methods, however, the virtual screening is still facing the issue of small sizes and insufficient structural diversity of the chemical compound libraries used in these endeavors.

Thus fragment-based approaches have been proposed to provide an alternative. Ludi is one example of computer-aided design [24] while SAR by NMR is an example of *in vitro* based design [25]. Both *in silico* and *in vitro* methods are rooted in a similar concept. One starts with identifying molecular fragments that can bind to potential pockets in molecular targets and then one links those fragments with appropriate linkage fragments to form a single molecule that has the highest binding affinity. This approach dramatically reduces the conformational space and increases the structural variety of molecules. However, combinatorial explosion is the bottleneck in this type of design.

*3.2. Drug Likeness*

3.2.1. Lipinski's Rule of Five

The compound design strategies discussed above do not guarantee that compounds discovered will become drug candidates. In order for a compound to be considered a valid drug candidate, it needs to possess drug-like properties. Several drug-like properties that are normally introduced in the drug discovery process as screening criteria are discussed below.

In 1997, Lipinski *et al.* [26] reported an intensive review of the physico-chemical criteria for orally administered active compounds that have poor solubility and permeability. They found that when a compound has a molecular weight (MWT) larger than 500, LogP greater than 5, its structure possesses more than five H-bond donors or more than 10 H-bond acceptors, the compound is highly likely to exhibit poor absorption. Namely, the compound is unlikely to be absorbed into the blood circulation system and reach active sites. This, so-called Lipinski rule of five has become a gold standard and a first filtering step in screening medicinal chemistry compounds. Namely, it helps to screen out compounds designed for oral administration that have a higher chance to be absorbed.

3.2.2. Pharmacokinetics (Absorption, Distribution, Metabolism, Excretion and Toxicity)

However, even though compounds may possess good absorption, it does not guarantee that such compounds can be considered "drugs". After absorption, as a drug, a molecule must be able to reach active sites that are believed to be the molecular sources causing the disease in question and bind to target proteins to regulate their functions with the highest efficacy. To achieve a desired efficacy, the amount of the drug entering active sites will need to be maintained at a specific level after this drug has undergone various types of biotransformation such as intestinal and liver metabolism. In the meantime, the toxicity caused by off-target effects needs to be reduced to a minimum. Although there are no quantitative criteria defined based on these pharmacokinetic properties, this provides a qualitative guideline to increase the probability that the screened compounds can achieve desired pharmacokinetic properties. With the advance of chemo-informatics and bioinformatics, ADMET predictor from Simulation Plus Inc. [27] is a relevant example of software that uses computational methods to predict absorption, distribution, metabolism, excretion and toxicity of given compounds.

3.2.3. Pharmacodynamics

The last criterion we wish to address is pharmacodynamics of compounds. It describes the response of the body to the drug and the mechanisms of drug action. As is well known, the goal of

designing a drug is to produce desired physiological responses in the body given a specific amount of the drug has been administered over a particular amount of time. Therefore, the knowledge of the pharmacodynamics of compounds certainly provides crucial information for screening out promising compounds as drugs. Unfortunately, one cannot directly conduct *in vivo* experiments without identifying lead compounds. One solution is to utilize *in vitro-in vivo* correlation (IVIVC) prediction, if such a correlation exists to predict *in vivo* effects from *in vitro* studies [28,29]. Searching a better way to establish IVIVC has become one of the most important subjects of study in the pharmaceutical industry to better design drugs and optimize their formulation.

## 4. Maximum Entropy in Drug Discovery

Previously, we have visited several key subjects and some commonly used methods in drug discovery. Now, we focus on the way maximum entropy can be introduced in drug discovery as either a tool or a reasoning framework for developing methods to solve problems of relevance to drug discovery. Specifically, we discuss three subjects: (a) target identification; (b) compound design and (c) pharmacokinetics and pharmacodynamics. Note that our intention is not to try and address the methodology used, their advantages and shortcomings and results obtained in the work leading to the discovery of new drugs. Instead, we wish to direct the reader's attention to the central logic behind these applications in terms of the entropy-based scheme. Namely, what is the right question to ask to solve the problems involved in drug discovery. Once the right question is posed, the use of maximum entropy becomes a unique and straight-forward tool to answer it. Note that since most examples except aptamer design illustrated below are not entirely tackled through the entropy-based scheme we will use aptamer design example to quantitatively demonstrate the use of this scheme. We will qualitatively illustrate rest of examples in terms of this scheme.

### 4.1. Target Identification

As mentioned above, the systems biology approach gradually leads to the development of many analytical tools to evaluate the likelihood of biological targets being relevant drug targets. One of the key elements in these tools is defining the most appropriate way to interpret and characterize network data. For example, Yang *et al.* define the activity of a target based on the reaction rate to characterize the corresponding networks [19]. Therefore, potential drug targets can be selected according to the ranking order of the reaction rate.

Alternatively, Fuhrman *et al.* [30] proposed utilizing Shannon entropy, Equation (1), to quantify information codified in activities. Their goal is to identify potential drug targets through information changes in temporal targets' gene expression patterns, which are either genes that underwent a disease process or a normal phenotypic change.

The rationale behind their method is based in the meaning of entropy and in the methods to codify information into a probability distribution $p_i$ that is relevant to the gene expression level. They use entropy $S[p] = -\sum_i p_i \log p_i$ as a measure of variation in a series of gene expression data that describe a certain biological processes. Therefore, for genes with no variation in the expression level, they assign zero entropy and hence they are considered to contain no information. Namely, these genes are

not responsible for triggering the processes in question. Next, they define probability distributions through histograms of expression levels that are believed to contain information relevant to the changes of gene expression levels due to these processes. The key to successfully quantifying information content in temporal genes expression patterns hinges on an appropriate definition of the biologically meaningful expression level (namely, the bin size in a histogram).

In short, the logic behind the method of Fuhrman *et al.* [30] becomes even much clearer in terms of the entropy-based scheme. The whole idea of identifying drug targets is to ask "*What is the probability distribution of occurrence of a specific gene expression level that is relevant to the disease process of interest?*"

### 4.2. Compound Design

### 4.2.1. Docking

In the framework of information theory, the first and third criteria mentioned in the section on the design strategy above can be considered to represent the fundamental information required in virtual screening approaches. The second criterion then can be treated as an information processing guideline. Therefore, the efficiency and accuracy of this step will depend on the methods chosen for information processing. The genetic algorithm is but one of such examples. The algorithm borrows the concept of a genome evolution process to search conformations of a complex of targets and chemical compounds.

Chang *et al.* [31] have shown a better alternative, namely MEDock. Although MEDock did not completely exploit the entropy-based scheme for conformational searches, it still utilizes the maximum entropy principle as a guideline to make decisions in this search process. The fundamental question asked in MEDock is "*What is the probability of finding the deepest energy valley in the ligand-target interaction energy landscape*?" Maximum entropy suggests a direction to update the initial guess of a binding mode (described by an almost uniform distribution) to an optimal mode (a localized distribution around the global energy minimum).

### 4.2.2. Aptamer Design

The second example for compound design discusses utilizing the entropy-based scheme to design aptamers. Aptamers are short nucleic acid sequences typically identified through an experimental technique, the Systematic Evolution of Ligands by EXponential enrichment (SELEX) [32,33]. Aptamers can bind to specific molecular targets including small molecules, proteins, nucleic acids, phospholipids and can also be targeted to complex structures such as cells, tissues, bacteria and other organisms. Because of their strong and specific binding through molecular recognition, aptamers are promising tools in molecular biology and have both therapeutic and diagnostic clinical applications [32–35]. Unfortunately, some limitations of the SELEX technique have seriously slowed down the progress of discovering aptamers [36].

With the help of entropy-based inductive inference, a fragment-based approach has been developed to "design" aptamers given the structure of the target of interest [36]. The concept behind the fragment-based approach is to ask the question: "*Given the structural information about the target, what is the preferred probability distribution of having an aptamer that is most likely to interact with*

*the target*?" A solution to this problem was found using an entropy-based inductive inference with the seed-and-grow strategy. The solution involves acquisition of two types of information, what is the relevant information for determining the first nucleotide, seed, and given the preferred nucleotide what is the relevant information for determining the next one. The approach first determines the preferred probability distribution of the first nucleotide $P_{\mathrm{nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right)$ that likely interacts with the target. Since one information that is relevant to target and first nucleotide interaction is total energy of this complex, it is reasonable to consider expected total energy of target denoted by "tar" and first nucleotide denoted by "nt1",

$$\left\langle H_{\mathrm{Total}}\left(Q_{\{i\}}, q_{\{j\}}^1\right)\right\rangle = \sum_{Q_{\{i\}}, q_{\{j\}}^1} P_{\mathrm{nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right) H_{\mathrm{Total}}\left(Q_{\{i\}}, q_{\{j\}}^1\right), \tag{4}$$

as preliminary information in hand, where Hamiltonian of target-first nucleotide complex is $H_{\mathrm{Total}}\left(Q_{\{i\}}, q_{\{j\}}^1\right) = H_{\mathrm{tar}}\left(Q_{\{i\}}\right) + H_{\mathrm{nt1}}\left(q_{\{j\}}^1\right) + H_{\mathrm{tar\text{-}nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right) + H_{\mathrm{solvent}}$ and $Q_{\{i\}} = \left\{R_{\{i\}}, P_{\{i\}}\right\}$ stands for $\{i = 1, ..., N_Q\}$, *i.e.*, the atomic spatial coordinates and momenta of the $N_Q$-atom target and $q_{\{j\}}^1 = \left\{r_{\{j\}}^1, p_{\{j\}}^1\right\}$ stands for the $\{j = 1, ..., N_{\mathrm{nt1}}\}$ atomic spatial coordinates and momenta of the first $N_{\mathrm{nt1}}$-atom nucleotide. Furthermore, $H_{\mathrm{tar}}\left(Q_{\{i\}}\right)$ is the total internal energy of the target "tar", $H_{\mathrm{nt1}}\left(q_{\{j\}}^1\right)$ is the total internal energy of the first nucleotide "nt1", $H_{\mathrm{tar\text{-}nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right)$ denotes the interaction energy of the target and the first nucleotide and $H_{\mathrm{solvent}}$ represents the solvation energy.

Afterward, the approach keeps updating the probability distribution when more nucleotides are added to the previous one. Again, the total energy of the target and all nucleotides (previous and the added one) provides a reasonable choice as new information relevant to interactions of targets and additional nucleotides. For example, information acquired in the second step is the expected total energy of the target and two-mer nucleotides:

$$\left\langle H_{\mathrm{Total}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right)\right\rangle = \sum_{Q_{\{i\}}, q_{\{j\}}^1} P_{\mathrm{nt2}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right) H_{\mathrm{Total}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right), \tag{5}$$

where $q_{\{j\}}^2 = \left\{r_{\{j\}}^2, p_{\{j\}}^2\right\}$ stands for the $\{j = 1, ..., N_{\mathrm{nt2}}\}$ atomic spatial coordinates and momenta of the second $N_{\mathrm{nt2}}$-atom nucleotide. Maximizing relative entropy:

$$S\left[P_{\mathrm{nt2}} | P_{\mathrm{nt1}}\right] = -\sum_{Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2} P_{\mathrm{nt2}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right) \log \frac{P_{\mathrm{nt2}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right)}{P_{\mathrm{nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right)} \tag{6}$$

then determining probability distribution of two-mer nucleotides interacting with the target, $P_{\mathrm{nt2}}\left(Q_{\{i\}}, q_{\{j\}}^1, q_{\{j\}}^2\right)$, which is updated from $P_{\mathrm{nt1}}\left(Q_{\{i\}}, q_{\{j\}}^1\right)$. Continually using Equation (3), maximum entropy allows us to determine to what extent this update is sufficient and a sequence of nucleotides that is likely to bind to the target can be determined. The method has been applied to design aptamers which have been computationally and experimentally validated to bind specifically to targets such as thrombin [36], phosphatidylserine [37] and galectin-3 (still under experimental confirmation).

*4.3. Pharmacokinetics and Pharmacodynamics*

As mentioned earlier, there are many subjects regarding prediction of pharmacokinetic and pharmacodynamic properties of lead compounds. We will only discuss two examples here to demonstrate the use of entropy-based scheme.

4.3.1. Drug absorption Rate Estimation

The absorption of drugs involves several processes including dissolution, gastric and intestinal emptying and plasma disposition. Several methods have been developed for predicting drug absorption. One traditional way is to utilize statistical approaches to deduce models from training sets for absorption prediction [38]. Their success relies on the selection of an appropriate training set and statistical approaches. Another approach is based on a mechanistic model such as the well-known mass balance approach [39], a multi-compartment model [40] and an integration of these two [41]. This approach requires only *in vitro* acquired knowledge of drug permeability through the gut membrane and diffusion properties without conducting any clinical studies. However, the complexity of the absorption process introduces many uncertainties in linking permeability and the absorption rate.

At present, to the best of our knowledge, there is surprisingly only one maximum entropy method developed to estimate drug absorption rates based on the knowledge of the drug's disposition kinetics and plasma concentration profiles obtained from subjects who receive drugs intravenously and orally [42]. Although this method was designed to analyze plasma concentration data and lacks a predictive power because of the need for clinical data, it may still shed important light on developing a mechanistic model that is more appropriate than a mass balance theory-based approach commonly used now.

The crux in Charter and Gull's approach [42] is inductive inference based on the Bayes theorem. The question asked is "What is the probability distribution of a drug being absorbed given experimental plasma concentration profile data and mechanistic processes such as diffusion relevant to the absorption process?" Charter and Gull [42] argued that utilizing inductive inference provides a correct mechanism to tackle uncertainty in the absorption process and noise in the data. Furthermore, because the method incorporates mechanistic kinetics relevant to the absorption process, there is no need for additional preprocessing of experimental data for analysis and an estimate of the absorption rate.

Although Charter and Gull's method [42] was designed to analyze experimental data to estimate the absorption rate, we argue that one may apply the same approach to develop probabilistic kinetics and replace traditional deterministic kinetics. The method introduced in Usansky and Sinko [41] may provide a way to integrate probabilistic kinetics into multi-compartment models. Therefore, it may lead to a more appropriate description of complex absorption processes.

4.3.2. Prediction of the Response to Multidrug Combinations

In the last example presented here, we discuss the use of maximum entropy in the prediction of pharmacodynamics of multidrug combinations. As mentioned earlier, the multiple-drug multiple-target paradigm gradually becomes to be the trend in modern drug discovery. However, statistical mechanics has clearly demonstrated that many-body problems are intractable without introducing approximations

because of complicated many-body interactions. When one attempts to investigate synergistic and antagonistic effects of a mixture of multiple drugs through drug-drug and drug-target interactions, it is inevitable to face the same difficulty as in statistical mechanics descriptions of many-body systems.

Following empirical knowledge, Chou [43] proposed a generic approximation method to study synergism and antagonism for mixtures of multiple drugs. The idea is to approximate complicated many-body interactions by single-drug effects and pair-wise interactions. Furthermore, mechanistic equations such as the mass-action law and the equilibrium law are fundamental relations used to describe dose-response relations. Alternatively, Wood *et al.* [44] utilized an entropy scheme to develop a mechanism-independent method for predicting effects of mixtures of multiple drugs.

Similar to Chou's method [43], a pair-wise interaction approximation is also considered in the Wood *et al.* method. Furthermore, the crux of the method is to ask the question "*What is the joint probability distribution of a single drug and pair interactions given observed responses of single drugs and pairs of drugs*?" Based on this probability distribution, one can then estimate the expected response for multiple drugs. Since the parameters required in the probability distribution are determined through the use of empirical data, there is no need for any mechanistic relations and this avoids intractable problems due to many-body interactions. However, the main shortcoming is that these parameters do not lend themselves to any obvious physiological interpretation. Yet this method still provides information for inferring relationships between the response and pair-wise interactions. Although the method was originally proposed to investigate the effects of multidrug combinations in bacterial growth, the entropic framework is independent of biological systems. Namely, the same framework can be applied to systems other than bacteria.

## 5. Conclusions

We have shown several applications of entropy-based schemes ranging from the identification of biological targets to the prediction of pharmacodynamics of drugs in the field of drug discovery. Our discussions emphasize an important concept illustrated by the work of Caticha [11]. Caticha argues that entropy has many faces including the degree of randomness, information measure and is also a tool for inductive inference. This is all correct and merely depends on how one wants to use it. There is no need to seek profound interpretations. Knowing "how" to use entropy is the key. It will assist us to reason, explore and develop theories to solve problems of interests.

As we have addressed in each subject dealt with in this paper, the key to solve problems in the framework of an entropy-based scheme is to ask the right question first. And then one needs to seek out information relevant to the problem. Thereafter, the principle of maximum entropy leads to the most honest inference one can make to solve the problem given the limited amount of information about the system. Although the topics in drug discovery discussed here were tackled using different aspects of entropy, we argue that there actually exists a unified reasoning framework behind all of these applications. Namely, one can say that maximum entropy likely is the central concept that governs the choice of the right drug targets, the structures of proper regulators and appropriate combinations of them that lead to optimal pharmacokinetics and pharmacodynamics. We can expect to see more applications of this entropy-based scheme in the future. For example, one can apply a similar strategy used in aptamer design to develop a fragment-based approach to design small molecules to

bind to specific targets based on information such as the one relevant to interactions between chemical fragments (functional groups) and targets and the ways of linking these fragments. In pharmacodynamics, the entropy-based scheme will definitely provide a least biased approach to analyze and model *in vitro* cell response studies, in which one sometimes cannot obtain sufficient data due to biological or experimental limitations. This is exactly the problem that entropy-based scheme is designed to tackle. Namely, one is trying to answer what is the most preferred model one can build given insufficient cellular response data. Hopefully, this unified reasoning framework may in the future attract more attention and lead to the establishment of such an information-driven foundation in drug discovery.

## Acknowledgments

## Author Contributions

The authors contributed equally to the presented mathematical and computational framework and the writing of the paper.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Caticha, A. Entropic Inference and the Foundations of Physics; Monograph Commissioned by the 11th Brazilian Meeting on Bayesian Statistics—EBEB-2012; USP Press: São Paulo, Brazil, 2012; Available online: http://www.albany.edu/physics/ACaticha-EIFP-book.pdf (accessed on 8 March 2012).
2. Cox, R.T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **1946**, *14*, 1–13.
3. Cox, R.T. *The Algebra of Probable Inference*; The Johns Hopkins Press: Baltimore, MD, USA, 1961.
4. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
5. Jaynes, E.T. Information theory and statistical mechanics II. *Phys. Rev.* **1957**, *108*, 171–190.
6. Caticha, A. Entropic priors. In Proceedings of the 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole, WY, USA, 3–8 August 2003; Erickson, G., Zhai, Y., Eds.; AIP: Melville, NY, USA, 2004; Volume 707, pp. 371–380.
7. Shore, J.E.; Johnson, R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37.
8. Shore, J.E.; Johnson, R.W. Properties of cross-entropy minimization. *IEEE Trans. Inf. Theory* **1981**, *27*, 472–482.

9. Giffin, A.; Caticha, A. Updating probabilities with data and moments. In Proceedings of the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 8–13 July 2007; Knuth, K.H., Caticha, A., Center, J.L., Giffin, A., Rodríguez, C.C., Eds.; AIP: Melville, NY, USA, 2007; Volume 954, pp. 74–84.

10. Caticha, A. Information and Entropy. In Proceedings of the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 8–13 July 2007; Knuth, K.H., Caticha, A., Center, J.L., Giffin, A., Rodríguez, C.C., Eds.; AIP: Melville, NY, USA, 2007; Volume 954, pp. 11–22.

11. Caticha, A. Relative Entropy and Inductive Inference. In Proceedings of the 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole, WY, USA, 3–8 August 2003; Erickson, G., Zhai, Y., Eds.; AIP: Melville, NY, USA, 2004; Volume 707, pp. 75–96.

12. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Soc.: Providence, RI, USA, 2000.

13. Rodríguez, C.C. Entropic Priors for Discrete Probabilistic Networks and for Mixtures of Gaussian Models. In Proceedings of the 22nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering; Moscow, Idaho, 3–7 August 2002; Fry, R.L., Ed.; AIP: Melville, NY, USA; Volume 617, pp. 410–432.

14. Editorial office. Bigger isn't always better. *Nature* **2002**, *418*, 353.

15. Food and Drug Administration. *Innovation and Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*; FDA White Paper, 2004. Available online: http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/ CriticalPAthOpportunitiesReports/ucm113411.pdf (accessed on 28 March 2014).

16. Sams-Dodd, F. Target-based drug discovery: Is something wrong? *Drug Discov. Today* **2005**, *10*, 139–147.

17. Szuromi, P.; Vinson, V.; Marshall, E. Rethinking drug discovery. *Science* **2004**, *303*, 1795.

18. Csermely, P.; Agoston, V.; Pongor, S. The efficiency of multi-target drugs: The network approach might help drug design. *Trends Pharmcol. Sci*. **2005**, *26*, 178–182.

19. Yang, K.; Bai, H.; Ouyang, Q.; Lai, L.; Tang, C. Finding multiple target optimal intervention in diseaserelated molecular network. *Mol. Syst. Biol*. **2008**, *4*, 228.

20. Martí-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sánchez, R.; Melo, F.; Šali, A. Comparative Protein Structure Modeling Of Genes And Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–324.

21. Honma, T. Recent advances in de novo design strategy for practical lead identification. *Med. Res. Rev*. **2003**, *23*, 606–632.

22. Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649–663.

23. Morris, G.M.; Huey, R.; 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**, *16*, 2785–91.

24. Bohm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **1992**, *6*, 61–78.

25. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

26. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

27. Simulations Plus Inc. *ADMET Predictor*$^{TM}$ *7*; Simulations Plus Inc.: Lancaster, CA, USA, 2014.

28. Leeson, L.J. In vitro/In vivo correlations. *Drug Inf. J.* **1995**, *29*, 903–915.

29. US Department of Health and Human Services. Guidance for industry: Immediate release solid oral dosage forms scale-up and postapproval changes: Chemistry, manufacturing, and controls, *in vitro* dissolution testing, and *in vivo* bioequivalence documentation. Center for drug Evaluation and Research. 1995; Available online: http://www.fda.gov/downloads/Drugs/Guidances/UCM070636.pdf (accessed on 28 March 2014).

30. Fuhrman, S.; Cunningham, M.J.; Wen, X.; Zweigerm, G.; Seilhamer, J.J.; Somogyi, R. The application of Shannon entropy in the identification of putative drug targets. *Biosystems* **2000**, *55*, 5–14.

31. Chang, D.T.-H.; Oyang, Y.-J.; Lin, J.-H. MEDock: A Web Server For Efficient Prediction of Ligand Binding Sites Based on A Novel Optimization Algorithm. *Nucleic Acids Res.* **2005**, *33*, W233–W238.

32. James, W. Aptamer. In *Encyclopedia of Analytical Chemistry*; Wiley & Sons Inc.: Hoboken, NJ, USA, 2000; pp. 4848–4871.

33. Nimjee, S.M.; Rusconi, C.P.; Sullenger, B.A. Aptamers: An emerging class of therapeutics. *Annu. Rev. Med.* **2005**, *56*, 555–583.

34. Hamula, C.L.A.; Guthrie, J.W.; Zhang, H.; Li, X.F.; Chris, L.X. Selection and analytical applications of aptamers. *Trends Anal. Chem.* **2006**, *25*, 681–691.

35. James, W. Aptamers in the virologists' toolkit. *J. Gen. Virol.* **2007**, *88*, 351–364.

36. Tseng, C.-Y.; Ashrafuzzaman, M.D.; Mane, J.Y.; Kapty, J.; Mercer, J.R.; Tuszynski, J.A. Entropic fragment based approach for aptamer design. *Chem. Biol. Drug Des.* **2011**, *78*, 1–13.

37. Ashrafuzzaman, M.D.; Tseng, C.-Y.; Kapty, J.; Mercer, J.R.; Tuszynski, J.A. A Computationally Designed DNA Aptamer Template with Specific Binding to Phosphatidylserine. *Nucleci Acids Ther.* **2013**, *23*, 418–426.

38. Zhao, Y.H.; Le, J.; Abraham, M.H.; Hersey, A.; Eddershaw, P.J.; Luscombe, C.N.; Butina, D.; Beck, G.; Sherborne, B.; Cooper, I.; *et al.* Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* **2001**, *90*, 749–784.

39. Sinko, P.J.; Leesman, G.D.; Amidon, G.L. Predicting Fraction Dose Absorbed in Humans Using a Macroscopic Mass Balance Approach. *Pharm. Res.* **1991**, *8*, 979–988.

40. Yu, L.X.; Amidon, G.L. A compartmental absorption and transit model for estimating oral drug absorption. *Int. J. Pharm.* **1999**, *186*, 119–125.

41. Usansky, H.H.; Sinkko, P.J. Estimating Human Drug Oral Absorption Kinetics from Caco-2 Permeability Using an Absorption-Disposition Model: Model Development and Evaluation and Derivation of Analytical Solutions for $k_a$ and $F_a$. *J. Pharmacol. Exp. Ther.* **2005**, *314*, 391–399.

42. Charter, M.K.; Gull, S.F. Maximum Entropy and Drug Absorption. *J Pharmacokinet. Biopharm.* **1991**, *19*, 497–520.

43. Chou, T.-C. Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacol. Rev*. **2006**, *58*, 621–681.

44. Wood, K.; Nishida, S.; Sontag, E.D.; Cluzel, P. Mechanism-independent method for predicting response to multidrug combinations in bacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12254–12259.