

COMPUTATIONAL FRAMEWORK FOR RATIONAL DESIGN OF ALLOSTERIC DRUG CANDIDATES

Technical Field

The present invention relates, in general terms, to a computational framework for rational design of allosteric drug candidates.

Background

In the discovery of new medicine, it is crucial to control functions of target macromolecules through a binding of small molecule compounds (i.e. ligands) to a primary site of function (i.e. functional sites). Functional sites consist of protein residues performing or regulating a catalytic activity or interactions with other molecules or macromolecules. Conventionally, most current drugs inhibit or promote a function upon binding to functional sites. The mode of action of these drugs are commonly described as orthosteric.

There are several challenges typical for the development of orthosteric drugs. As functional sites on a target protein may also present in homologous proteins (i.e. proteins having similar structure and sequence of functional sites), there is a possibility of the orthosteric drugs/ligands binding to off-target proteins. This may lead to toxicity and other unwanted side effects.

Allostery is a pervasive phenomenon, where a protein's function is modulated by remote perturbations, such as ligand binding and mutations, which occur at a distance from the primary functional sites. In a corresponding approach to drug discovery, there has been work done on designing allosteric drugs (also termed generically as effectors, i.e. ligand working allosterically) that remotely modulate a protein function upon binding to locations outside of functional sites. These sites are generally referred to as allosteric sites.

In the allosteric mode of action, binding of allosteric effectors/drugs to allosteric sites causes alteration in the protein dynamics, which, in turn, produce long-range effects at remote protein sites. The allosteric communication between two distant sites can be characterized by the energetics of allosteric signalling, which describe changes in the structure and/or dynamics at an affected site upon a perturbation at distant allosteric site. Noteworthy, the allosteric signalling is reversible. Conformational changes at an allosterically affected functional site may cause gradual modulation of the protein function, which may lead to a beneficial outcome in various therapeutic applications.

The advantages of using allosteric modes of action in drug development, disease diagnostics, and biological engineering of therapeutic biologics are well recognised. To name the major ones: (i) allosteric effectors generally possess high selectivity and specificity, thus allowing to alleviate drug toxicity; (ii) the allosteric effects are modulatory rather than by the way of on-off, allowing fine-tuned control of protein function.

Present challenges surrounding drug design of allosteric drug candidates include the identification of latent allosteric sites and prospective effectors capable of modulating a protein's function.

Moreover, current approaches for uncovering allosteric sites and effectors typically focus on identifying druggable binding sites with desirable geometric (such as non-flat pocket-like surfaces) and stereochemical properties. While these approaches generally work for the discovery of traditional orthosteric drugs that bind to functional site pockets with mostly conserved features, allosteric sites are highly diverse and specific to different proteins. Moreover, allosteric sites may contain structure/sequence-wise featureless and/or flat surfaces. Therefore, it is unlikely that employing knowledge derived from orthosteric drugs/sites in the tasks of identifying latent allosteric sites and designing new allosteric drugs will be successful.

Another critical issue in the development of allosteric effectors/drugs is that current approaches – mostly based on large-scale screening using chemical libraries, typically disregard the role of allosteric signalling, which lies at the core of the problem. In particular, questions, such as what is the allosteric signalling required to modulate the protein function? What remote perturbations (e.g. binding, mutations, and their

combinations) may originate the desired allosteric signalling? How allosteric signalling may be originated by the structural dynamics of the protein? All the above questions have not been formulated and addressed so far.

There is now a need for exploring and understanding the most salient features of allosteric binding sites, corresponding effectors, as well as an explicit quantification of allosteric signalling from allosteric to functional sites and its tuning. The presented computational framework provides a foundation for the rational design and engineering of allosteric drug candidates capable of specific and targeted modulation of a protein's function, by considering the protein structure and dynamics.

Summary

Disclosed herein is a method of identifying at least one of allosteric drug candidates, their binding sites, site-effector interactions, allosteric signalling to one or more target functional sites, and optimisation of site-effector pairs by adjusting or tuning of site-effector composition and interactions, the method comprising the following steps:

- i. obtaining, for a protein of interest, 3D structure information and information on a protein function to be a target of allosteric modulation;
- ii. identifying residues, of the protein of interest, that modulate one or more target functional sites of the protein of interest, by perturbing every residue or three-residue segment and quantifying allosteric signalling from each residue or three-residue segment to all other residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
- iii. identifying residues that modulate one or more target functional sites, by perturbing a functional site and quantifying allosteric signalling to all residues, at single-residue resolution in the protein of interest, based on the 3D structure information;

- iv. identifying a set of binding patches associated with residues that are allosterically linked to at least one of the one or more target functional sites as identified in steps ii and iii, and selecting one or more binding patches of interest from the set of binding patches by quantifying allosteric signalling from each binding patch to a target functional site, geometrical and chemical properties of the respective patch;
- v. for each binding patch of interest, identifying one or more ligands and corresponding binding sites;
- vi. either:
 - determining a set of allosteric site-effector pairs, each allosteric site-effector pair comprising a said ligand and a corresponding binding site as identified in step v, based on allosteric signalling from each pair to a said target functional site, the binding free energy, and physicochemical properties of the pair; or
 - if the one or more ligands do not possess a desired allosteric signalling and binding free energy, selecting a further one or more binding patches of interest and repeating from step v, and/or adjusting the compounds in step v, and repeating from step vi.
- vii. outputting the set of site-effector pairs.

In some embodiments, identifying one or more compounds comprises screening or designing a plurality of compounds for the binding patches, wherein the one or more compounds are selected from the plurality of compounds.

In some embodiments, the method in step vi comprises ranking compounds in the plurality of compounds based on allosteric modulation and predicted binding free energy, and identifying the one or more compounds based on the ranking.

In some embodiments, identifying the one or more compounds comprises identifying one or more compounds that possess a desired allosteric effect and binding free energy.

In some embodiments, obtaining the 3D structure information comprises performing one or more of X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryogenic electron microscopy, or performing computational structural prediction.

In some embodiments, step ii involves simulating binding of a probe, using Structure-Based Statistical Mechanical Model of Allostery (SBSMMA), the model accounting for causality of allosteric communication and quantifying energetics of allosteric signalling and modulation, to each three-residue segment in a protein chain of the protein of interest, and/or, separately, simulating mutations at all residue positions, and calculating the allosteric modulation to all residues.

In some embodiments, step iii involves Reverse Perturbation (RP) to identify allosterically-linked protein residues by simulated binding at one of the target functional sites. The RP method is based on a concept that allosteric communication is reversible – simulated perturbations such as ligand binding at a regulated site can originate allosteric signalling to regulatory sites (hence "Reverse"). As a result, latent allosteric sites or new potential binding sites allosterically linked to regulated functional ones can be identified.

In some embodiments, the processes of screening or designing compounds for binding patches and of ranking compounds based on allosteric modulation value and predicted binding affinity are repeated for two or more rounds.

In some embodiments, during each round except for a final round, an analogue search and optimization process of obtained compounds is performed and an output of the process serves as inputs for a subsequent round. The final round outputs compounds, which are accepted, and no further search is performed.

In some embodiments, steps ii, iii, iv, and vi involve use of Structure-Based Statistical Mechanical Model of Allostery (SBSMMA), the model accounting for causality of allosteric communication and quantifying energetics of allosteric signalling and modulation.

In some embodiments, quantifying energetics of allosteric signalling and modulation involves an analysis of protein dynamics, an evaluation of per-residue allosteric

potential, and a computation of the per-residue free energy associated with allosteric signalling based on a protein conformational ensemble of the protein of interest.

In some embodiments, step vi comprises ranking the one or more site-effector pairs based on allosteric modulation, predicted binding free energy to the binding sites and physicochemical properties of site-effector pairs.

In some embodiments, step vi comprises searching corresponding analogues and derivatives of the one or more compounds based on substructure matching or similarity and the ranking.

In some embodiments, step iv comprises evaluating and prioritising the one or more binding patches based on allosteric modulation and a composition of residues.

In some embodiments, the one or more binding patches are evaluated and prioritised based on properties of known allosteric sites in proteins and/or properties of known allosteric compounds for proteins.

In some embodiments, the one or more compounds are evaluated and prioritised based on molecular weight, shape and aromaticity.

In some embodiments, the one or more compounds are evaluated and prioritised based on properties of known allosteric sites in proteins and/or properties of known allosteric compounds for proteins.

In some embodiments, for at least one of the one or more compounds, step vi comprises conducting a search of a chemical space surrounding the at least one compound, to identify further candidates of allosteric effectors based on substructure matching or similarity.

In some embodiments, the one or more allosteric site-effector pairs are evaluated, prioritised and adjusted based on energetics of allosteric signalling and modulation at the one or more target functional sites, binding affinity and/or properties of site-effector compositions and interactions.

In some embodiments, a simultaneous and mutual design of effector-site pairs with desirable properties comprises mutual adjustment of binding sites and effectors, via fine-tuning of the composition of allosteric site-effector pairs and intermolecular interactions of site-effector pairs, in respective adjustment, evaluation and prioritization steps across successive rounds.

In some embodiments, evaluation, prioritisation and simultaneous and mutual adjustment of a site-effector pair continues in iterations until a site-effector pair with desired allosteric modulation and binding affinity is obtained.

The term "desirable" refers to properties, allosteric modulation, physicochemical properties, binding free energy and so on, complying with predetermined requirements. The predetermined requirements may refer to, for example, modulation or binding free energy being at or above a predetermined level, or within a predetermined range, or comprising specific physicochemical properties. As used herein, the term "effector" refers to a ligand that can cause allosteric signalling.

Disclosed herein is also a system for identifying at least one of allosteric drug candidates, their binding sites, site-effector interactions, allosteric signalling to one or more target functional sites, and optimisation of site-effector pairs by adjusting or tuning of site-effector composition and interactions, the system comprising:

a processor;

a memory device, accessible to the processor, the memory device comprising program code executable by the processor to:

- i. obtain, for a protein of interest, 3D structure information and information on a protein function to be a target of allosteric modulation;
- ii. identify residues, of the protein of interest, that modulate one or more target functional sites of the protein of interest, by perturbing every residue or three-residue segment and quantifying allosteric signalling from each residue or three-residue segment to all other residues, at

single-residue resolution in the protein of interest, based on the 3D structure information;

- iii. identify residues that modulate one or more target functional sites, by perturbing a functional site and quantifying allosteric signalling to all other residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
- iv. identify a set of binding patches associated with residues that are allosterically linked to at least one of the one or more target functional sites as identified in steps ii and iii, and select one or more binding patches of interest from the set of binding patches by quantifying allosteric signalling from each binding patch to a target functional site, geometrical and chemical properties of the respective patch;
- v. for each binding patch of interest, identify one or more ligands and corresponding binding sites;
- vi. either:
 - determine a set of allosteric site-effector pairs, each allosteric site-effector pair comprising a said ligand and a corresponding binding site as identified in step v,, based on allosteric signalling from each pair to a said target functional site, the binding free energy, and physicochemical properties of the pair; or
 - if the one or more ligands do not possess a desired allosteric signalling and binding free energy, select a further one or more binding patches of interest and repeating from step v, and/or adjusting the compounds in step v, and repeat from step vi.
- vii. output the set of site-effector pairs.

As used herein, the term "patch" refers to an initial location on a protein (among many other locations) that is analysed to determine its properties – e.g., allosteric modulation. Similarly, the term "site" refers to a location having a set of protein residues that interact with a ligand revealed by simulation discussed below.

Brief description of the drawings

Some embodiments of the systems and methods for this invention, in accordance with the present disclosure, will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Figure 1 illustrates key conceptual distinctions between orthosteric ligands and sites and allosteric effectors and sites. Similar orthosteric ligands are obtained for a conserved site involved in the protein function. On the other hand, diverse effectors are obtained for various distinct allosteric sites with different structures and properties, binding of which send allosteric signals to modulate the distant functional site.

Figure 2 provides a schematic diagram of major modules of the computational framework for generating allosteric drug candidates. The Allosteric Site Identification Module takes a protein structure as an input and identifies potential allosteric sites which may contain at least one residue that is allosterically linked to a functional site of interest. A protein residue is allosterically linked to a functional site if a perturbation of the former (residue) originates an allosteric signal (with desirable allosteric modulation, see below) to the latter (functional site), or vice versa. The module uses the Reverse Perturbation method to identify residues allosterically linked, i.e. capable of sending an allosteric signal to a functional site upon perturbation, to a target site. The protein structure also serves as an input to compute Allosteric Fingerprints, which contain complete data on allosteric modulation caused by each perturbation. Targeted Analysis and/or Agnostic Analysis are performed on the Allosteric Fingerprints to identify specific perturbations that can cause the desired allosteric modulation at target site and/or multiple protein regions, respectively. An allosteric modulation exerted to a site, in terms of the energetics of signalling Δh evaluated in the framework, is considered desirable if the magnitude $|\Delta h|$ is larger than the energetics of thermal fluctuation at 0.6 kcal/mol and the sign of the resulted modulation corresponds to the allosteric effects required, according to specifics of the task ($\Delta h > 0$ indicates conformational changes and $\Delta h < 0$ indicates stabilization). Results from the Reverse Perturbation and Allosteric Fingerprints delineate multiple binding patches on the protein that can result in required modulation upon binding of an effector to it. The Chemical Data Module applies statistics and knowledge on general properties of known allosteric sites and effectors

for obtaining large sets of existing or new compounds/biologics from chemical libraries or design methods. In the Evaluation Module, binding of multiple compounds/biologics to multiple binding patches is evaluated. Results and properties of corresponding effectors and their bound sites are assessed, and the most promising initial candidates are selected. Optionally, the initial effectors and sites can provide feedback to adjust the binding patch generation and selection step and/or the selection and design of compounds/biologics, to yield optimized allosteric site-effector pairs.

Figure 3 provides a flow diagram of the computational framework to generate and design allosteric drug candidates in accordance with various embodiments.

Figure 4 depicts a conceptual illustration of a computational processing system capable for implementing all aspects of the system in Figure 3.

Figure 5 provides data showing the allosteric effects of known agonists (DEL, KGC and DPI) and antagonists (NH2 and NAL) binding to the endogenous pocket of the delta opioid receptor, in terms of per-residue allosteric modulation (kcal/mol) at each residue of the entire structure and at each residue constituting the GAL site (indicated by open circles), computed by using SBSMMA.

Figure 6 shows the results from applying the Reverse Perturbation method on a structure of the delta opioid receptor. Left: perturbation of protein residues constituting the known functional GAL site. Right: per-residue allosteric modulation caused by perturbation at the distant GAL site overlaid on the structure.

Figure 7 shows the Allosteric Signalling Fingerprint and Allosteric Probing Fingerprint obtained for a structure of the delta opioid receptor.

Figure 8 provides data showing shortlisted distant residues that are allosterically coupled to the GAL site identified using the Reverse Perturbation method. The data also show shortlisted distant probes and mutations that can cause the largest allosteric modulation at the GAL site, identified by performing the Targeted Analysis and Agnostic Analysis on the Allosteric Probing Fingerprint (probes) and Allosteric Signalling Fingerprint (mutations).

Figure 9 shows designated binding patches 1, 2 and 3 consisting of shortlisted residues obtained from the Reverse Perturbation, Targeted Analysis and Agnostic Analysis methods. Binding to patch 1 is likely to originate positive allosteric modulation in the distant GAL site, whereas binding to patches 2 and 3 is likely to cause negative modulation in the same site.

Figure 10 provides data showing property distributions of allosteric and orthosteric drugs/drug candidates and sites. (a) Ratio of volume versus surface area of a drug; (b, c) numbers of aliphatic rings and aromatic rings normalized by the molecular weight; (d) molecular weight of the drug; (e, f) numbers of hydrogen bond donors and acceptors normalized by the drug's molecular weight; (g) average number of neighbouring residues for residues in a given site; (h, i) numbers of apolar and polar residues; (j, k, l) numbers of negatively, positively charged and aromatic residues.

Figure 11 provides data showing an evaluation of fragment-like compounds obtained by applying the DSiP fragment library on binding patch 1 on the delta opioid receptor. Left: A plot of allosteric modulation (kcal/mol) and binding free energy (kcal/mol) of shortlisted docked ligand poses obtained from a molecular docking technique. The allosteric modulation is computed by applying SBSMMA, using protein residues constituting the binding site of a docked ligand and the protein structure as inputs. The binding free energy is computed by the performed molecular docking. A fragment-like compound D-S1-1 is indicated by a filled circle on the plot. Right: the structure of selected compound D-S1-1 and its binding conformation on patch 1.

Figure 12 shows property distributions of compounds derived by substructure matching using the fragment-like compound D-S1-1 as a reference. Quantitative Estimate of Drug-likeness (QED), number of aromatic ring per molecular weight (AROM/MW), the second normalized principal moments of inertia (NPR2) and a distribution of the NPR2 versus NPR1 are presented. Multiple derived compounds were shortlisted for further evaluation based on criteria indicated as dashed lines on the charts.

Figure 13 provides data showing an evaluation of allosteric modulation and binding free energy of a set of shortlisted compounds. For instance, six derived compounds (D-S1-2 to D-S1-7) and the initial fragment-like compound (D-S1-1) causing large

binding free energy and allosteric modulation are indicated by filled circles. The chemical structures of derived compounds are depicted.

Figure 14 provides charts that show binding energy, allosteric modulation, QED, AROM/ MW and NPR2 of D-S1-1 to D-S1-7. Depending on the tasks and requirements for designing allosteric drugs, different prospective candidates can be selected.

Figure 15 illustrates binding conformations of D-S1-2 to D-S1-7 and their interactions with residues constituting the binding site (defined as residues located within 4Å from any heavy atom of a bound ligand).

Figure 16 provides data showing evaluation of allosteric modulation and binding free energy for D-S2-1 to D-S2-7 obtained for patch 2, and D-S3-1 to D-S3-7 obtained for patch 3. These compounds are instances of allosteric effectors that bind to distant locations and originate negative allosteric modulation to the GAL site. The chemical structures of derived compounds and initial fragment-like compound are depicted.

Figure 17 provides charts that show binding energy, allosteric modulation, QED, AROM/ MW, NPR2 and logP of D-S2-1 to D-S2-7 for patch 2, and D-S3-1 to D-S3-7 for patch 3. Depending on the tasks and requirements for designing allosteric drugs, different prospective candidates can be selected.

Figure 18 shows the Allosteric Signalling Fingerprint and Allosteric Probing Fingerprint obtained for a structure of the oncogenic K-Ras^{G12V}.

Figure 19 illustrates binding patches 1 and 2 obtained from the Agnostic Analysis performed on the Allosteric Fingerprints. Using the patch residues and the protein structure as inputs, SBSMMA showed that binding to patch 1 causes larger allosteric modulation at distant locations, including SW1 and SW2 sites, compared to patch 2.

Figure 20 provides data showing an evaluation of compounds obtained by applying the DCM library to binding patch 1 on the K-Ras^{G12V}. Left: To identify compounds that cause a large and specific allosteric modulation at the SW1 site only, the modulation at SW1 site is subtracted by the modulation at SW2 during the evaluation of bound DCM compounds. As an instance, compound K-S1-1 which binds to patch 1 and originates large allosteric modulation to SW1 is indicated by a filled circle on the plot.

Right: the chemical structure of selected compound K-S1-1 and its binding conformation on patch 1.

Figure 21 shows the chemical structures of K-S1-1 and its molecular fragments obtained from the RECAP (Retrosynthetic Combinatorial Analysis Procedure) fragmentation method. Further compounds are obtained by substructure matching using fragment 2 as a reference.

Figure 22 shows property distributions of compounds derived from fragment 2 including QED, AROM/MW, NPR2 and NPR2 versus NPR1. Multiple derived compounds were shortlisted for further evaluation based on criteria indicated as dashed lines on the charts.

Figure 23 provides data showing an evaluation of derived compounds in terms of allosteric modulation to SW1 and binding free energy to patch 1 on K-Ras^{G12V}. Left: six derived compounds (K-S1-2 to K-S1-7) as instances that cause large binding free energy and allosteric modulation, and the reference DCM compound (K-S1-1), are indicated by filled circles. Right: the chemical structures of K-S1-1 to K-S1-7 are depicted.

Figure 24 provides charts that show binding energy, allosteric modulation, QED, AROM/ MW and NPR2 of K-S1-1 to K-S1-7. Depending on the tasks and requirements for designing allosteric drugs, different prospective candidates can be selected.

Figure 25 illustrates intermolecular interactions of derived compounds K-S1-2 to K-S1-7 with binding site residues in patch 1. The varied site-effector pairs shown here result in different binding free energy and allosteric modulation at the distant SW1 site.

Detailed description

The present invention describes a generic computational framework to predict allosteric sites on any protein and to perform rational design of effectors that may be utilised as prospective drug candidates. Embodiments of the present invention relate

to systems or methods for identification of allosteric sites and design of allosteric drug candidates by means of computation. In particular, embodiments incorporate computational modelling and analysis on proteins and compounds, which can be further extended to any soluble/globular protein regulated allosterically and any compound that can work as an allosteric effector.

A generated chemical structure of an effector may be utilised to synthesise a small molecule effector or a macromolecule, including but not limited to peptides, polypeptide chains, protein domains and biologics. Synthesised effectors may be assessed via experimental and computational means for their ability to bind residues in designated allosteric sites and/or modulate a function of its associated protein and/or producing biological effects that may be beneficial for therapeutic applications. Various therapeutic applications include, but are not limited to, treatments, prevention, and diagnostics or relief of symptoms related to a single or multiple diseases.

The present invention uses the Structure-Based Statistical Mechanical Model of Allostery (SBSMMA). SBSMMA may do two things. It may quantify the energetics of allosteric signalling and account for the causality of allosteric communication. This enables allosteric signalling strength to be determined and used as a basis for identifying binding patches that are useable to achieve the desired therapeutic application, and selecting binding patches with highest allosteric signalling strength. It may also be utilised for locating latent allosteric sites and for design of allosteric effectors. SBSMMA has been benchmarked on large and heterogeneous sets of allosteric proteins. SBSMMA has been implemented on the publicly available Allosteric Signalling and Mutation Analysis (AlloSigMA) web server and results of calculations performed with SBSMMA are stored in the Allosteric Mutation Analysis and Polymorphism of Signalling (AlloMAPS) database.

SBSMMA comprises three components. The first component involves an analysis of protein dynamics as inputs, wherein the protein dynamics may be obtained via computing the normal modes from elastic network models of the protein or covariance matrices from molecular dynamics simulations of the protein. The second component involves an evaluation of per-residue allosteric potential, and the third involves obtaining the free energy of allosteric signalling (kcal/mol) due to a perturbation. Explicit quantification of the energetics of purely allosteric effects, termed here as allosteric modulation (kcal/mol), is obtained from the free energy

associated with an allosteric signalling. The allosteric modulation describes a causality of long-range communication and predicts triggering of potential conformational changes in the protein structure and dynamics caused by perturbations. The per-residue free energy of allosteric signalling may be computed based on conformational ensembles of the protein of interest, which describe the dynamics of neighbouring residues.

Figure 1 illustrates key conceptual distinctions between orthosteric ligands and sites and allosteric effectors and sites. Similar orthosteric ligands are obtained for a conserved site involved in the protein function. On the other hand, diverse effectors are obtained for various distinct allosteric sites with different structures and properties, binding of which sends allosteric signals to modulate a distant functional site.

A main distinction between allosteric and orthosteric binding sites is that allosteric binding sites may be designated on any location separated from a target functional site, whereas orthosteric binding sites overlap (at least partially) with a target functional site. Consequently, a multiplicity of allosteric binding sites generally exhibits greater diversity in terms of structure and stereochemistry than one or a few orthosteric sites. Figure 1 illustrates the said diversity. Correspondingly, effectors typically show more diverse variations in terms of chemical structures and molecular properties compared to orthosteric ligands. Design of prospective effectors is generally less restrictive as effectors are not required to possess similar structure and/or molecular properties with respect to endogenous ligands or macromolecules that bind the functional site to be modulated, as opposed to orthosteric ligands.

Figure 2 provides a conceptual illustration of major processes in the framework, which consists of three modules – Allosteric Site Identification Module, Chemical Data Module and Evaluation Module. The Allosteric Site Identification Module identifies a plurality of overlapping or non-overlapping binding patches that may originate desired allosteric modulation to target sites or regions. Importantly, the identification of allosteric binding patches follows a process of rational design and tuning, which may involve an exhaustive computation of the free energy associated with allosteric signalling on a single-residue resolution and other methods. The Chemical Data Module applies statistics and knowledge of known allosteric effectors and sites to select and shortlist a plurality of existing compounds or biologics from chemical

libraries or generate new ones via *de novo* molecular design or fragment-based design approaches. In the Evaluation Module, specifics of intermolecular interactions between bound allosteric sites and associated effectors, and their compositions, may serve as inputs for evaluating and determining key determinants of site-effector pairs yielding desired properties, including the allosteric modulation and binding free energy. This information may be provided to the next rational design-evaluation cycle, in an iterative process for continuous optimization and fine-tuning of the site-effector pairs. Whether a property is "desired" will depend on the protein being investigated, and the specifics of the task (e.g., to promote or inhibit a particular function of the protein) that the compound being discovered is intended to achieve. Desired properties may include, but are not limited to, allosteric signalling to a functional site, or geometric and chemical properties of binding patches. A shortlisted binding patch of interest should at least be able to send a stronger allosteric signal (with large magnitudes of quantified allosteric modulation), compared to those not being selected, to the targeted functional site. Selection involves quantifying or determining the presence of the properties mentioned above and then, on the basis of that quantification or determination, selecting binding patches. For example, binding patches may be selected where allosteric signalling for those patches is higher than a predetermined threshold, is in the top X% of allosteric signalling strength of the patches being considered for selection, or among the top *N* patches being considered for selection in terms of allosteric signalling strength.

The framework yields a set of site-effector pairs as final outputs for allosteric modulation of a target functional site of a protein (Figure 2). The binding free energy of a site-effector pair and produced allosteric modulation are determined by its composition and the site-effector interactions. A composition of a site-effector pair (i.e. site-effector composition) comprises the chemical structure of the effector and interacting protein residues in the binding site, providing information at the atomistic resolution on spatial arrangement, geometry and intramolecular covalent bonds of respective effector and interacting protein residues. Site-effector interactions comprise atomistic information on intermolecular non-covalent and covalent interactions/bonds between an effector and residues in its binding site, including but not limited to the type of interactions/bonds (e.g. hydrogen bonds, salt bridges and disulfide bonds), atoms involved and geometry of said intermolecular interactions.

Fine-tuning and optimization of the allosteric modulation, binding free energy and physicochemical properties of site-effector pairs comprise concomitant adjustments of the corresponding composition and intermolecular interactions of said pairs – this may be referred to as "adjustment", "fine tuning" or similar, of the compound or effector or of the binding patches or sites. An adjustment of the chemical structure of an effector, via identification or design of a plurality of effectors and selecting one or more of those effectors over the other effectors in the plurality of effectors, may result in corresponding changes in the set of interacting site atoms and residues. Conversely, adjusting a composition of atoms and residues constituting a desired binding site may differentially select for distinct effectors capable of binding thereto. This is described as mutual adjustments of sites and effectors (site-effector pairs), wherein an adjustment of a binding site affects an identification or design of effectors, and vice versa.

In the traditional orthosteric drug design, candidate compounds are identified based on their specific interactions with a small number of residues that are key to the protein function. In rational design of allosteric drug candidates, varied allosteric signalling may be induced or originated by effector binding to a large repertoire of sites, each containing a distinct set of binding residues and a bound effector (Figure 1). The induced allosteric signalling may be fine-tuned upon iterative mutual adjustments of site-effector pairs. Therefore, the rational design and mutual adjustments of a plurality of site-effector pairs with highly diverse compositions and intermolecular interactions is an important advantage of this presented framework.

The mutual adjustments of identified allosteric sites (in the Allosteric Site Identification Module) and designed effectors (in the Chemical Data Module) to yield new site-effector pairs are performed based on information feedback from the Evaluation Module, where the allosteric modulation, binding free energy and physicochemical properties of previously obtained site-effector pairs are evaluated in a previous cycle.

Figure 3 provides a detailed flow diagram depicting methods and operations in the computational framework for rational design of allosteric drug candidates.

Allosteric Site Identification Module

At operation **302**, a three-dimensional structure of a target protein is provided as an input to the module. Advantageously, the present invention has a minimal requirement for execution of the framework, the single requirement being the target protein's structure. Knowledge of the target protein's structure allows all necessary data to be produced upon execution of the framework. Obtaining a protein structure comprises performing one or more of experimental structural determination methods, such as X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryogenic electron microscopy or computational structural prediction methods. Computational structural prediction methods may include AlphaFold and RoseTTAFold.

At operation **304**, the user may choose to provide a list of residues defining the composition of a target functional site via a user interface. The target site residues may be designated based on previous knowledge on the protein function and/or requirements of the task.

In the case of a defined functional site, the Reverse Perturbation (RP) method is performed (method **306**) to identify distant residues that are allosterically linked to the functional site. The RP method is based on a concept that allosteric communication is reversible – perturbations at a known functional site may originate allosteric signalling to its regulatory sites (hence "Reverse"). In the RP method, binding of a ligand to a defined functional site is performed or simulated on a protein structure, followed by a measurement of allosteric signalling to all other residues in terms of energetics, from which a small number of residues that are most strongly modulated by the perturbations are selected. According to the RP approach, a residue is allosterically linked to a perturbed functional site if the residue shows the largest magnitudes of, or strongest, allosteric modulation (kcal/mol), when compared with other residues (e.g., top 10% in magnitude, or top *N* residues), due to a perturbation at the functional site. In the presented framework, up to 40 and at least 10 residues that show the largest magnitudes of modulation are used. The number of selected residues may be varied and should depend on the total number of residues in a target protein. A group of proximal (e.g., located within a predetermined distance of each other) allosterically linked residues form an allosterically linked protein region which may contain latent allosteric sites that target the functional site. The RP method may be executed using the SBSMMA.

The Allosteric Fingerprints method (method **308**) includes computing the Allosteric Signalling Fingerprint (ASF) and Allosteric Probing Fingerprint (APF) based on a protein structure. The ASF and APF are computed to obtain exhaustive information on the energetics of allosteric signalling between all positions at a single-residue resolution, due to mutations and binding of small fragment-like compounds, respectively. In ASF, two kinds of mutations are simulated iteratively at each residue position – UP mutation where the residue is replaced by a bulkier one; DOWN mutation where the residue is replaced by a smaller one. The energetics of signalling is measured in terms of the allosteric modulation (kcal/mol) upon each mutation. Obtaining the difference of modulation caused by UP and DOWN mutations at each residue position, known as the modulation range, provides a generic description of a maximal strength of allosteric signalling caused by perturbing a residue position. The exhaustive ASF contains all modulation range values from every position to all other positions in a protein structure. In APF, binding of a probe to every consecutive three-residue segment is simulated iteratively and the allosteric modulation (kcal/mol) exerted at all other residue positions are measured. The Allosteric Fingerprints method may be executed using the SBSMMA.

Advantageously, due to the extensivity of the free energy, the exhaustive Allosteric Fingerprints may be utilised to explore and adjust combinations of per-residue perturbations (e.g. mutations and/or probe binding) in various design methods of allosteric effectors and corresponding binding sites, including but not limited to fragment-based design.

The Reverse Perturbation and Allosteric Fingerprints (**306** and **308**) are complementary methods that may be performed in tandem as complementary components. The former rapidly identifies residues that may constitute an allosteric site using only information of a regulated functional site. The latter provides a complete description and quantification of all possible signalling in a protein structure at a single-residue resolution, allowing to directly identify residues that specifically originate desired signalling to targeted positions or broadly affecting large regions of a structure.

The Targeted Analysis (method **310**) identifies residue positions that can result in the strongest allosteric modulation at one or more defined functional sites upon a mutation or binding of a small molecule with a low molecular weight (also termed as

a probe here) , using the ASF and APF as data source, respectively. The Agnostic Analysis (method **312**) identifies residue positions that can result in the strongest averaged allosteric modulation to all distant positions upon a mutation or probe binding, using the ASF and APF as data source, respectively. Some of these positions identified by method **312** may constitute a functional site. In the case where information on a functional site is available, methods **310** and/or **312** may be performed after obtaining the ASF and APF. Alternatively, if such information is not initially available, method **312** may be performed.

At operation **314**, residue positions shortlisted from the RP and Targeted/Agnostic Analysis (methods **306**, **310** and **312**) are used to generate a plurality of binding patches on the protein structure. A binding patch defines a location on a structure in which effectors may be evaluated for their binding affinities, allosteric modulations and other properties. A binding patch may constitute a group of residues that originate strong allosteric signalling as identified and shortlisted from one or more of methods **306**, **310** and **312**, and may be in any number and combination, provided that constituent residues are in proximity between each other. A binding patch may also contain residues that are not identified by methods **306**, **310** and **312**, but are in proximity with those identified by the methods. Various binding patches may contain overlapping groups or completely distinct groups of residues.

At operation **314**, computing allosteric modulation caused by binding to patch residues evaluates a plurality of binding patches generated to produce a smaller shortlisted set. Allosteric modulation may be computed using SBSMMA. Other selection criteria may include, but not limited to, geometrical and chemical properties of binding patches, such as the solvent-accessible surface area, and various properties of constituting residues and their combinations.

Advantageously, the present invention addresses the problems of allosteric site prediction and identification of binding effectors concurrently, allowing both tasks to mutually adjust and feedback to each other. Specifically, the plurality of binding patches generated, each capable of originating strong allosteric modulation and possessing varied properties, may also be selected in relation to properties of shortlisted compounds or biologics, produced by operation **318** in the Chemical Data Module.

Chemical Data Module

It is generally recognised that allosteric and orthosteric sites and ligands may exhibit or require different properties due to their different modes of actions. At operation **316**, statistical data on general properties of a plurality of known allosteric sites and effectors may be collected and analysed. A comparison of the above with those obtained from known orthosteric sites and ligands may also be performed to identify key properties that show distinctions in the characteristics of allosteric sites and ligands/effectors versus orthosteric sites and ligands, this information may be used to facilitate or ensure identification or design of binding sites or effectors with allosteric characteristics. The data and results may be used as an input for subsequent steps.

At operation **318**, one or more chemical libraries may be selected for screening if they contain existing compounds or biologics with properties based on or in correspondence with properties of known allosteric effectors revealed by operation **316**.

Alternatively, computer-aided molecular design, including but not limited to *de novo* generative modelling and fragment-based design, may be performed to obtain new compounds or biologics at operation **318**. Design methods may be selected if they generate compounds or biologics with properties based on or in correspondence with properties of known allosteric effectors revealed by operation **316**.

At operation **318**, one or more chemical libraries or design methods may be selected based on or in correspondence to properties of known allosteric sites, and as such may exhibit strong binding free energy (indicating high binding affinity). The binding free energy may be computed using molecular docking methods, such as AutoDock Vina. For instance, if operation **316** indicates known allosteric sites contain a higher proportion of hydrophobic residues compared to orthosteric sites, compound libraries containing existing compounds or design methods capable of generating new compounds with high aromatic ring contents may be selected.

At operation **318**, one or more chemical libraries or design methods may be selected based on or in correspondence to properties of binding patches shortlisted by operation **314** to originate strong allosteric modulation. Advantageously, the mutual

adjustment and feedback between Allosteric Site Identification and Chemical Data modules facilitate identification and design of allosteric drug candidates that may bind strongly to allosteric sites and, at the same time, elicit required allosteric signalling upon binding.

At operation **318**, one or more chemical libraries or design methods may be selected if they contain compounds or biologics (compounds and biologics being ligands in the present context) with properties that may be beneficial in various aspects associated with drug efficacy and toxicity. For instance, chemical libraries with high selectivity may be utilised as the compounds or biologics are generally expected to cause less off-target toxicity/effects.

At operation **318**, a selection of one or more chemical libraries or design methods may also be guided by other considerations, including but not limited to specific requirements on properties of desirable drug candidates and various aspects associated with follow-up chemical synthesis, experimental testing/validation, exploration of the structure-activity relationships, and lead optimization.

At operation **320**, a plurality of compounds or biologics from selected chemical libraries or design methods may be further shortlisted based on their properties, including but not limited to molecular weights, shapes, quantitative measures of drug-likeness, and a presence/absence of certain functional groups. This shortlisting step may be advantageous if selected libraries or design methods yield large numbers of compounds or biologics.

After performing operation **320**, a plurality of shortlisted compounds or biologics with diverse chemical structures and varied properties may be obtained. Together with shortlisted binding patches from operation **314**, they may serve as inputs to the Evaluation Module.

Evaluation Module and Mutual Adjustment of Site-Effector Pairs

At operation **322** in the Evaluation Module, one or more shortlisted binding patches (**314**) may serve as locations near which intermolecular interactions between protein residues therein and a plurality of diverse existing/new compounds or biologics (shortlisted by operation **320**) are predicted, simulated, and explored. Conformations

and interactions of a plurality of compounds or biologics to residues in binding patches may be obtained by methods including, but not limited to molecular docking.

Diverse compounds or biologics may interact with varied overlapping or non-overlapping groups of residues within a binding patch, consequently exhibiting different properties and/or causing various effects. Therefore, before characterizing each intermolecular interaction, a binding site may be defined as a plurality of protein residues that are in proximity to the bound compound/biologic (also termed as a ligand). For instance, a binding site may consist of residues with at least one heavy atom separated from at least one heavy atom of the associated ligand by a distance not more than a certain limit, such as 4 angstrom.

At operation **322**, the energetics of allosteric signalling caused by a ligand binding and measured in terms of allosteric modulation, may be computed by using the associated binding site and the protein structure as inputs for SBSMMA. Computing the allosteric modulation allows to identify binding sites that may be utilised as allosteric sites, and, thus, obtained ligands that may be considered and utilised as allosteric effectors.

At operation **322**, prediction and simulation of binding interactions between a plurality of compounds or biologics and a plurality of binding patches, produce as outputs a variety of corresponding pairs of allosteric sites and effectors, termed as "site-effector pairs". Potential allosteric drug candidates may be identified from a variety of site-effector pairs.

At operation **322**, properties associated with each site-effector pair may be recorded and evaluated, including but not limited to (i) chemical structure of the effector, which may be described by a spatial arrangement of its atoms and chemical bonds in two-dimensional and/or three-dimensional space; (ii) molecular descriptors of the effector; (iii) residue composition of the binding site; (iv) specifics of non-covalent interactions between the effector and site residues; (v) the free energy of binding (kcal/mol), which may indicate the binding affinity between the site-effector pair, and (vi) allosteric modulation (kcal/mol) caused by binding of the effector to the site residues, which may indicate changes in dynamics and structure at distant affected sites.

At operation **322**, site-effector pairs with very low binding free energies may be discarded from further analysis, as they are unlikely to form any stable interaction with the site. Subsequently, site-effector pairs may be analysed, ranked and selected according to their allosteric modulations and binding free energies, both are important quantitative measures for shortlisting potential allosteric drug candidates.

If selected top-ranking site-effector pairs exhibit desired allosteric modulations and binding free energies to similar extents, other properties, including but not limited to chemical structure and molecular descriptors, composition of binding site, and specifics of non-covalent interactions, may be used to further shortlist potential allosteric site-effector pairs.

At **324**, if it is determined that further refinement or optimization of site-effector pairs outputted is required, adjustments may be performed for the site-effector pairs in operation **326**.

Advantageously, operation **326** forms an important iterative feedback process of the present invention. Depending on site-effector data from **322** and **324**, it provides feedback to (i) operation **314** in the Allosteric Site Identification Module to adjust selection and shortlisting of binding patches; and (ii) operations **318** and **320** in the Chemical Data Module to adjust collection/generation, selection, and shortlisting of compounds/biologics. As a result, obtained binding patches and effectors are continuously modified, mutually adjusted, and evaluated to yield new and optimized site-effector pairs, in multiple iterations, until site-effector pairs with desired properties are obtained.

Advantageously, adjustment of binding patches may lead to corresponding changes in the screening or design of effectors, and vice versa. The mutual adjustment allows to address concurrently the optimization tasks of allosteric sites and allosteric effectors – the “site-effector pairs” – through iterative exploration, design and fine-tuning of site-effector interactions to produce desirable allosteric modulation, binding free energy and other properties.

An instance of effector adjustment at operation **326**, a derivation of new compounds may be achieved through an identification of structural analogues based on matching or similarity of chemical structures. Alternatively, smaller sub-structures may be

obtained via molecular fragmentation, larger compounds containing the sub-structures may be subsequently derived. The latter method is aligned with a fragment-based design approach, which may be used in the design methods for designing initial compounds from chemical libraries containing fragment-like compounds (**318**) and in the follow-up derivation of new compounds during the adjustment step (**326**). This method may allow for a more thorough coverage over the chemical space in the Chemical Data Module, in order to increase the rate at which a promising allosteric drug candidate is identified.

Comparative analysis may also be performed in addition to a derivation or modification of effectors aforementioned – by comparing various properties of high-ranking and low-ranking ones in a previous search, distinctive improved effectors may be designed for subsequent iterations.

Derived or modified effectors may be evaluated and shortlisted based on their properties, including but not limited to molecular weights, shapes, quantitative measures of drug-likeness, and a presence/absence of certain functional groups. This shortlisting step may be advantageous if selected libraries or design methods yield large numbers of effectors.

Derived or modified effectors may be also evaluated and shortlisted based on or in correspondence with properties of known allosteric effectors and sites obtained at operation **316**, including aromaticity, molecular shape, and others.

An instance of allosteric site adjustment at operation **326**, a comparison of site-effector pairs with regard to outputted allosteric modulation and binding free energy in a previous iteration may identify site-effector interactions, at the residual or atomistic resolution that leads to desired allosteric modulation and binding free energy. Such information may guide the adjustment of binding patches generation and selection in the Allosteric Site Identification Module.

Advantageously, a key feature of the present invention, rapid computation of the energetics of allosteric signalling using the computationally efficient SBSMMA permits high-throughput evaluation during multiple iterations comprising screening, design and optimization of a plurality of site-effector pairs. The computation of the allosteric modulation delineates causality of the allosteric communication and quantifies it in

terms of allosteric free energies, showing a modulation of distant protein dynamics and function upon a variety of intermolecular site-effector interactions as perturbations.

At operation **328**, data and properties associated with final optimized site-effector pairs may be recorded, including but not limited to chemical structure and molecular descriptors of the effectors, residue composition of the binding site, specifics of non-covalent interactions, the binding free energy, and allosteric modulation of the site-effector pairs. Here, the site-effector pairs having the desired binding affinity and allosteric signalling may be outputted.

Computational Processing System

Figure **4** provides an illustration of a computational processing system for the rational design framework for allosteric drug candidates. In some embodiments, the computational processing system is contained within a single or multiple connected computing devices. In some embodiments, the system is implemented as a software to be executed on a computing or multiple connected computing devices.

A computational processing system includes: (i) an input/output interface for a user to submit a protein structure and/or designate a target functional site; (ii) a processor system including one or more of a CPU, GPU and/or other processing devices; (iii) a memory system storing data on chemical structures and housing computational methods or algorithms for molecular design; methods for executing the Allosteric Site Identification Module, Chemical Data Module and Evaluation Module; and storage for data on site-effector pairs and other outputs at the completion of the workflow.

Example 1: Design of Allosteric Drug Candidates for the Human Delta Opioid Receptor

The embodiments of the present invention will be better understood with the examples provided within. In the first example, the computational framework in the present invention is applied on the human delta opioid receptor (DOR).

The DOR belongs to the class A of G protein-coupled receptors (GPCRs), a superfamily of transmembrane protein receptors that forms a major drug target. The receptor is

chiefly involved in pain modulation and its activators are associated with potent therapeutic potential as analgesic drugs, antidepressants, and others. DOR activation is promoted by the binding of endogenous opioid peptides and other developed agonists (i.e. activators of GPCR function). Binding of an agonist at the endogenous pocket on the extracellular side may cause conformational changes in the distant intracellular site that promote interactions with the G alpha subunit (GAL site) via allosteric signalling, which, in turn, activate downstream intracellular signal transduction pathways.

Chief challenges associated with the development of drugs for DOR include the needs to achieve high selectivity to avoid off-target side effects due to unintended binding to other related opioid receptors. As DOR is involved in multiple biological pathways, another key issue is related to obtaining high activation specificity to targeted downstream intracellular pathways.

Prospective allosteric drugs may provide advantages in drugging DOR. For instance, binding to allosteric sites, which typically exhibit higher sequence diversity among homologous opioid receptors compared to that of the endogenous pocket, may allow to tune receptor selectivity. Moreover, the multiplicity and variety of all potential allosteric site-effector interactions may originate specific allosteric signalling to the GAL site for precise modulation of the DOR function (illustrated in Figure 5).

The computational framework in the present invention utilises a protein structure as an input. A structure of DOR is obtained using the AlphaFold protein structure prediction methods. Residues 40 to 335 forming the transmembrane domain of DOR are used. Figure 5 provides data of resulted allosteric modulation (kcal/mol) at all individual residues upon binding of different agonists and antagonists (i.e. inhibitors of GPCR function) at the endogenous pocket, demonstrating that the SBSMMA accounts for the allosteric signalling in DOR. A positive value of allosteric modulation may indicate conformational changes, whereas a negative value may indicate over-stabilisation of the structure. Notably, binding of agonists (Deltorphan II – “DEL”; KGCHM07 – “KGC”; DPI-287 – “DPI”) generally causes larger magnitudes of allosteric modulation compared to those by antagonists (DIPP-NH2 – “NH2”; Naltrindole – “NAL”). Moreover, binding of DEL, a natural agonist causes a larger modulation in DOR compared to the others. Specifically, DEL binding exerts strong positive modulation values at all GAL site residues, pointing to resulted conformational

changes at the distant GAL site that may, in turn, promote the binding of G alpha subunit and activation of downstream pathways. The binding sites of DEL, KGC, DPI, NH2 and NAL are defined for residues within 4 angstrom from respective peptide/ligands based on the Protein Data Bank structures (PDB ID: 8F7S, 6PT2, 6PT3, 4RWA and 4N6H). The GAL functional site is defined based on a PDB structure (8F7S).

After a demonstration of SBSMMA on the delta opioid receptor, the computational framework in the present invention is applied to identify latent allosteric sites and obtain new allosteric drug candidates that may originate different allosteric signalling to specifically modulate DOR function.

In the Allosteric Site Identification Module, the Reverse Perturbation (RP) method is performed to identify protein residues that are allosterically linked to the GAL site residues (Figure 6). Simulating a binding perturbation to GAL leads to positive modulation in the distant extracellular side of the receptor, as well as the intracellular H8 helix, pointing to a presence of latent allosteric sites in these locations.

Another method in the module, the Allosteric Fingerprints method is used to obtain complete information on all possible allosteric signalling in a protein structure (Figure 7) on a single-residue resolution. The Allosteric Signalling Fingerprint (ASF) and Allosteric Probing Fingerprint (APF) are computed for the DOR structure, providing data on allosteric modulation caused by mutations at every residue position and by the binding of a probe at every residue triplet, respectively. The ASF/APF data reveal a variety of positions that may be allosterically linked to protein regions and sites including the GAL functional site. According to the ASF and APF approaches, a residue is allosterically linked to a functional site if the residue causes the largest magnitudes of allosteric modulation (kcal/mol) at the functional site upon perturbing (e.g. mutations or binding of a probe) the residue. In this example, the Allosteric Fingerprints are computed using the SBSMMA.

The Targeted Analysis and Agnostic Analysis are performed on the ASF and APF to obtain residue positions that may constitute binding patches originating large positive, and separately, negative allosteric modulation targeting the GAL site. In the calculation of modulation exerted on a protein site, a mean value is obtained by averaging the modulation value of each residue within the site. The modulation value

(kcal/mol), which evaluates the free energy associated with the allosteric signalling on a single residue caused by a perturbation, may be compared with the energy of thermal fluctuations $k_B T$ (k_B : Boltzmann constant; T : temperature) at about 0.6 kcal/mol at room temperature. While some perturbations may appear to elicit a modulation lower than $k_B T$ at each functional site residue, on average; combinations of low-value modulations in a homogeneously affected site may still produce significant modulation of protein dynamics.

Targeted Analysis on the APF yields a list of 40 probes causing positive modulation values from 0.26 to 0.66 kcal/mol at the GAL site, and a lower number of probes causing negative modulation values at about -0.10 kcal/mol (Figure 8). Performing on the ASF, the analysis outputs a series of 40 residues causing positive modulation ranges from 1.10 to 2.05 kcal/mol upon mutations at the GAL site, and a lower number of residues causing negative modulation ranges not stronger than -0.42 kcal/mol.

Agnostic Analysis on the APF yields a list of 40 probes, the top-ranking probe binding causes 0.68 and -0.54 kcal/mol, in modulation value, at all positively and negatively modulated distant positions, respectively (Figure 8). Performing on the ASF, the top-ranking mutation causes 1.43 and -1.39 kcal/mol, in modulation ranges, at all positively and negatively modulated distant positions, respectively.

Based on the result from the RP method, a list of 40 positions that exhibit modulation values ranging from 1.11 to 2.10 kcal/mol is also collected.

Importantly, these complementary methods concurrently identified overlapping sets of residues within a protein region located on the extracellular side of DOR that may constitute binding patches for prospective allosteric drug candidates. In principle, a plurality of binding patches with different properties can be yielded and selected by varying and fine-tuning the residue composition. Depending on specifics of the tasks and/or inputs from other operations/methods in the framework, further shortlisting of binding patches may be performed.

Multiple binding patches with various properties, likely yielding a plurality of allosteric effectors with distinct properties and allosteric effects, are shortlisted to demonstrate the generic nature of the computational framework. For instance, residues 200-204

and 284-287, which are simultaneously outputted by the ASF, APF and RP methods are used to build a binding patch 1 (Figure 9). Targeted and Agnostic Analyses show that perturbations at these residues cause large positive values of allosteric modulation at multiple locations including the GAL site. In another instance, binding patch 2 is built from residues 267-269 and 313, which originate large negative values of allosteric modulation at the GAL site, as identified by the Targeted Analysis of ASF. In yet another instance, binding patch 3 is built from probes that originate large negative values of allosteric modulation at the GAL site (residues 225-228 and 268), as identified by the Targeted Analysis of APF. Binding patches 2 and 3 are proximal to each other, and both are located at the transmembrane portion of DOR, which interacts with the hydrophobic part of the plasma membrane. The shortlisted binding patches 1-3 are produced as outputs of the Allosteric Site Identification Module.

In the Chemical Data Module, data and statistics of known allosteric/orthosteric sites and drugs are collected and analysed (Figure 10). The analysis shows that effector drugs are generally flatter with a lower molecular volume to area ratio compared to orthosteric drugs. In another instance, effector drugs generally contain higher aromatic ring content, agreeing with the higher proportion of aromatic residues in known allosteric sites compared to those of orthosteric drugs/sites. Advantageously, the various distinctions between allosteric and orthosteric drugs and sites obtained in this module are utilised to select and shortlist prospective allosteric drug candidates in a knowledge-based approach, among a plurality of designed effectors obtained in multiple operations within the computational framework.

The DSI-P (Diamond-SGC-iNEXT Poised) library provides starting points for identification and fragment-based design of allosteric drug candidates for DOR. The DSI-P library contains small fragment-like compounds, with a library design principle that aims for rapid and cost-effective synthesis of "poised" fragments – fragments with at least one functional group that can be synthesized using a well characterized reaction.

In the Evaluation Module, the shortlisted binding patches are used as locations in which intermolecular interactions between protein residues therein and a set of 768 fragment-like compounds from the DSI-P library are predicted, simulated and explored. The conformations and intermolecular interactions of site-effector pairs are obtained by molecular docking using AutoDock Vina.

Figure 11 shows the evaluation results of site-effector pairs at binding patch 1 on DOR. For each DSiP compound, at least one and not more than ten bound conformations are sampled, each interacting with a different set of allosteric site residues. For each site-effector pair, the binding free energy characterizing the binding affinity is calculated using AutoDock Vina. Site-effector pairs with top 25% of binding free energy are analysed in further steps, as the rest are unlikely to form any stable interaction within the binding patch. For each of the 1902 site-effector pair yielded, allosteric modulation at the GAL site is computed using the high-throughput SBSMMA. Most site-effector pairs cause positive modulation values, indicating conformational changes at the GAL site upon the perturbations.

Advantageously, the evaluation step allows to identify or shortlist multiple site-effector pairs that may originate desired allosteric signalling to the GAL site while possessing sufficient affinity to the allosteric site.

As an example, a compound (D-S1-1), which shows large modulation and binding free energy is selected (Figure 11). Binding of D-S1-1 causes modulation at 1.05 kcal/mol and a binding free energy at -6.51 kcal/mol. Evaluation of its other properties including, for instance, the Quantitative Estimate of Drug-likeness (QED), number of aromatic ring per molecular weight (AROM/MW), the second normalized principal moments of inertia (NPR2) show values at 0.85, 8.29, and 0.89, respectively. The QED provides a continuous quantification of compound quality and potential as orally administered drugs, allowing to discard those with low likelihood of success in development. Furthermore, aromatic ring contents (AROM/MW) and molecular shape/flatness (NPR2) are also assessed, as statistical analysis in the Chemical Data Module shows that known allosteric effectors are typically flatter and contain more aromatic rings compared to the orthosteric ligands.

To further improve the various properties of site-effector pairs provided as output in the next iteration, D-S1-1 is used as an initial fragment for designing/deriving new compounds using a fragment-based design method. The large ZINC database containing more than 1 billion compounds is utilised to perform a substructure-based search in the chemical space, which identified 631 derivatives containing the D-S1-1 fragment as part of their chemical structures.

The Pan-Assay Interference Compounds (PAINS) filter is used to flag and remove derived compounds that may bind non-specifically and promiscuously to numerous biological targets, whereas the Brenk filter flags and removes derived compounds containing chemical moieties that are likely to be toxic and unstable. After applying both filters implemented in RDKit, 556 derivatives are obtained.

The properties of derived compounds are subsequently evaluated (Figure 12). For instance, QED, AROM/MW and NPR2 calculations are utilised to select derived compounds with properties that are generally consistent with those of known allosteric effectors. Other properties associated with the compounds may also be employed in a selection. 365 derivatives from D-S1-1 that simultaneously satisfy multiple criteria ($QED \geq 0.75$; $AROM/MW \geq 5$, $NPR2 \leq 0.95$) are shortlisted. Moreover, the NPR1/NPR2 distribution from the normalized principal moments of inertia analysis identified some derived compounds adopting a flat disc-like shape.

Binding of the shortlisted derivatives to binding patch 1 is simulated, followed by the computation and evaluation of the allosteric modulation and binding free energy of 1,803 varied site-effector pairs (using those with binding energies in the upper half of the distribution) derived from D-S1-1 (Figure 13). The operations in this iteration largely follow those in the previous iteration, a difference being the compounds evaluated.

Importantly, a plurality of newly derived compounds exhibits larger GAL modulation and/or binding free energy compared to D-S1-1. Several improved derivatives D-S1-2 to D-S1-7 are highlighted (Figure 13). The strong allosteric signalling observed may lead to conformational changes in the protein that produce large modulation of the protein function, and the stronger binding free energies indicate stability provided by the site-effector interactions. For comparison, the modulation and binding free energy of D-S1-1 (an instance of outputs from the previous iteration) are also indicated along with those evaluated in this iteration.

Figure 14 shows a comparison of various properties of D-S1-1 to D-S1-7, including the binding free energy, allosteric modulation, QED, AROM/MW and NPR2. Other properties of designed site-effector pairs may also be evaluated and form a basis for a selection. Depending on the requirements for prospective allosteric drug candidates and specifics of the design tasks, different effectors may be preferentially selected.

For instance, if effectors capable of originating strong allosteric signalling to the GAL site, possessing, at the same time, various molecular features that are highly desirable as orally administered drugs, D-S1-2 may be selected for its high modulation and QED values.

Corresponding pairs of effectors D-S1-2 to D-S1-7 and their binding site residues are visualized in Figure 15, showing the conservatism and diversity of site-effector pairs in various aspects, including but not limited to the residue composition of binding sites, conformation of effectors, and specifics of intermolecular interactions.

Designed allosteric drug candidates, including D-S1-2 to D-S1-7, may be outputted by the computational framework as final results, alternatively, they may be subject to further iterations consisting of substructure-search, evaluation and optimization to obtain other candidates, if needed.

The comparison and analysis of molecular and structural features in Figure 15 may complement the quantitative measurements of properties evaluated in Figure 14 in facilitating or guiding other operations in the framework, including but not limited to shortlisting of site-effector pairs and to providing inputs for design and/or adjustments of compounds and allosteric sites in the Chemical Data Module and Allosteric Site Identification Module.

The aforementioned operations and methods in the Evaluation Module are also applied on binding patches 2 and 3, which are built from residues originating negative allosteric modulation to the GAL site. Fragment-like compounds D-S2-1 and D-S3-1 from the DSi-P library are shortlisted for binding patches 2 and 3 in the first iteration. On the basis of these fragments, a plurality of effectors is designed/derived in the next iteration for binding patch 2 and 3. The evaluation results from the second iteration are presented in Figure 16, showing some derived effectors that exhibit negative allosteric modulation with a higher magnitudes and/or larger binding free energy, including the instances of D-S2-2 to D-S2-7 and D-S3-2 to D-S3-7, for binding patches 2 and 3, respectively. For comparison, the modulation and binding free energy of D-S2-1 and D-S3-1 (outputs from the previous iteration) are also indicated along with those evaluated in this iteration.

Importantly, the proximal binding patches 2 and 3 are exemplified here to illustrate that diverse site-effector pairs can be yielded in a general location, which in this case, being a part of the transmembrane domain of DOR surrounded by the hydrophobic portion of the plasma membrane. Diverse properties and outputs, including but not limited to the chemical structures of effectors, residue composition of allosteric sites, allosteric modulation, binding free energy, and other specifics of site-effector pairs can be obtained (Figure 16).

Moreover, depending on the requirements for prospective allosteric drug candidates and specifics of the design tasks, different binding patches and hence resulted site-effector pairs, may be preferentially selected. For instance, if effectors with high hydrophobicity (measured by the logP values) are desired, the effectors designed for binding patch 2 may be selected (Figure 17). In another instance, if effectors with high drug-likeness (measured by the QED values) are desired, the effectors designed for binding patch 3 may be selected.

Example 2: Design of Allosteric Drug Candidates for the Human K-Ras^{G12V} Oncoprotein

In the second example, the computational framework in the present invention is applied on the human K-Ras protein encoded by the KRAS (Kirsten rat sarcoma virus) gene. K-Ras is a monomeric GTPase with a catalytic domain highly conserved in the Ras family of GTPases. The protein plays critical roles in several signalling transduction pathways inside a cell, including the MAPK/ERK pathway involved in cellular proliferation and other biological processes.

The structure of the catalytic domain of K-Ras consists of two flexible regions known as the switch I (SW-I, residues 30-38) and switch II (residues 60-76), which play instrumental roles in various aspects of K-Ras regulation and function. Both SW-I and SW-II constitute the catalytic site, which performs GTP to GDP hydrolysis. The K-Ras function is tightly regulated by guanine nucleotide exchange factors and GTPase-activating proteins, which interact with SW-I and SW-II sites to facilitate GDP dissociation and GTP hydrolysis, respectively. Moreover, SW-I residues of activated K-Ras (GTP-bound state) are also involved in interacting and activating the function of various partner proteins in downstream pathways, resulting in a propagation of cellular signals to control various biological processes.

Various mutations of K-Ras are involved in multiple kinds of human cancers and other diseases including the G12V mutation. Pathogenesis associated with oncogenic K-Ras often involves a hyperactivated state of K-Ras, resulting in dysregulation of multiple pathways. Chief challenges surrounding the development of therapeutics for K-Ras oncoproteins include the high sequence conservation surrounding the catalytic site (i.e. the functional site) among Ras family of homologous proteins and GTPases in general. Hence, there is a need to achieve high selectivity to avoid off-target side effects due to unintended binding to other related GTPases. As the SW-I and SW-II sites of K-Ras are involved in multiple distinct processes, another key issue is related to obtaining high specificity for different downstream partner proteins, thus facilitating precise modulation of distinct biological pathways. The SW-I functional site is used as a target in this example.

A crystallographic structure of K-Ras^{G12V} is obtained from the PDB (PDB ID: 4TQ9) and serves as an input to identify latent allosteric sites, separated from the conserved functional site of K-Ras^{G12V}. Residues 1 to 168 forming the catalytic domain of K-Ras^{G12V} are used.

In the Allosteric Site Identification Module, the ASF and APF are computed for the K-Ras structure (Figure 18). The ASF/APF data reveal a variety of positions that may be allosterically linked to protein regions and sites, including the SW-I and SW-II functional sites.

In principle, the Targeted Analysis and Agnostic Analysis on the Allosteric Fingerprints, as well as the Reverse Perturbation methods, may be used together to designate allosteric binding patches. In this example, Agnostic Analysis on the ASF and APF is performed to identify residue positions that may constitute binding patches originating large allosteric modulation targeting SW-I and/or SW-II sites, and other distal locations. Starting from the "agnostic" binding patches, the identification and design of site-effector pairs with a SW-I specificity is subsequently explored to demonstrate flexibility and versatility of the generic framework in the present invention.

For instance, binding patch 1 (residues 94-98 and 136-138) and binding patch 2 (residues 118-122 and 141-143) are shortlisted from a list of patches built from

various combinations of residues obtained from the Agnostic Analysis (Figure 19). Calculation of per-residue modulation on the K-Ras^{G12V} structure using SBSMMA shows a binding perturbation at binding patch 1 causes a larger positive modulation compared to binding patch 2 at SW-I (0.49 versus 0.27 kcal/mol) and SW-II (0.32 vs 0.18 kcal/mol) sites, and other locations. Therefore, binding patch 1 is used in further steps of the framework.

In the Chemical Data Module, the DCM (Dark Chemical Matter) library is used to provide starting points for identification and design of allosteric drug candidates for K-Ras^{G12V}. The DCM library contains diverse compounds with novel chemotypes that are inactive in more than 100 different assays, indicating their potential as highly selective compounds and thus expected to exhibit low off-target toxicity, as opposed to promiscuous compounds, which often yield false positives in screening processes.

Utilising statistics and information on characteristics of known allosteric drugs/effectors, a set of criteria is used to select DCM compounds with properties that are generally aligned with known effectors. In principle, various combinations of properties may be used, including but not limited to those analysed in Figure 10. For instance, the following criteria are used: AROM/MW ≥ 4 , NPR2 ≤ 0.90 and molecular weight ≥ 350 Da, which identified 4,152 non-fragment-like and flat compounds with high aromatic ring contents out of 18,513 compounds from the DCM library.

In the Evaluation Module, binding patch 1 is used to explore and simulate intermolecular interactions between protein residues therein and 4,152 compounds selected from the DCM library. The operations and methods performed are largely similar to those executed in the first example on DOR. Following molecular docking, site-effector pairs with top 25% in binding free energy are analysed in subsequent steps. For each of the 10,348 site-effector pairs yielded, a measurement of differential allosteric modulation is computed by subtracting the SW-I modulation value by the SW-II modulation value. This measurement allows to obtain site-effector pairs capable of sending strong allosteric signal to SW-I, without affecting SW-II residues and its associated downstream pathways.

Figure 20 shows the evaluation results of site-effector pairs at binding patch 1 on K-Ras^{G12V}. Perturbations by site-effector pairs generally cause positive modulations at SW-I residues, indicating potential conformational changes in the functional site. As

an instance, a compound (K-S1-1), which shows large modulation at SW-I and negligible SW-II modulation, is selected. Binding of K-S1-1 causes modulation at 1.10 and 0.04 kcal/mol at SW-I and SW-II respectively, and a binding free energy at -6.23 kcal/mol. Evaluation of its other properties including, for instance, QED, AROM/MW, and NPR2 show values at 0.68, 4.81 and 0.87, respectively.

To improve the various properties of output site-effector pairs in the next iteration, K-S1-1 is used as an initial compound for designing/deriving new compounds. The RECAP (Retrosynthetic Combinatorial Analysis Procedure) molecular fragmentation method is applied on K-S1-1 to obtain constituent fragments that may serve as starting points for fragment-based design (Figure 21). As an instance, fragment 2 is used and 5,000 derived compounds are identified based on a substructure search performed on the ZINC library. Similar to Example 1, applying the PAINS and Brenk filters removed potentially toxic and non-selective compounds, and as a result yielded 3,485 derivatives for further steps.

The properties of derived compounds are subsequently evaluated (Figure 22). For instance, QED, AROM/MW, and NPR2 calculations are utilised to select derived compounds with properties that are generally consistent with those of known allosteric effectors. 849 derivatives from K-S1-1 that simultaneously satisfy multiple criteria ($QED \geq 0.75$; $AROM/MW \geq 5$, $NPR2 \leq 0.95$) are shortlisted. Moreover, the NPR1/NPR2 distribution from the normalized principal moments of inertia analysis identified some derived compounds adopting a flat disc-like shape.

In a new iteration, binding of the shortlisted derivatives to binding patch 1 is simulated, followed by the computation and evaluation of the allosteric modulation and binding free energy of 8,439 varied site-effector pairs (Figure 23). For comparison, the modulation and binding free energy of K-S1-1 (an instance of outputs from the previous iteration) are also indicated along with those evaluated in this iteration. While all derivatives show weaker SW-I modulation compared to K-S1-1, a plurality of derivatives with stronger binding free energy and other improved properties (such as QED) is obtained, including for instance derivatives K-S1-2 to K-S1-7.

Figure 24 shows a comparison of various properties of K-S1-1 to K-S1-7, including the binding free energy, allosteric modulation, QED, AROM/MW and NPR2. Other

properties of designed site-effector pairs may also be evaluated and form a basis for a selection. Depending on the requirements for prospective allosteric drug candidates and specifics of the design tasks, different effectors may be preferentially selected. For instance, if allosteric drug candidates capable of originating large and specific allosteric signalling to the SW-I site, possessing, at the same time, various "oral drug-like" molecular features indicated by the high QED value, the K-S1-2 may be selected.

Corresponding pairs of effectors K-S1-2 to K-S1-7 and their binding site residues are visualized in Figure 25, showing the conservatism and diversity of site effector pairs in various aspects, including but not limited to the residue composition of binding sites, conformation of effectors, and specifics of intermolecular interactions.

The designed allosteric drug candidates including K-S1-2 to K-S1-7 may be outputted by the computational framework as final results, alternatively, they may be subject to further iterations consisting of substructure-search, evaluation, and optimization to obtain other candidates, if needed.

Throughout this specification, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps, but not the exclusion of any other integer or step or group of integers or steps.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that the prior art forms part of the common general knowledge.

CLAIMS

1. A method of identifying at least one of allosteric drug candidates, their binding sites, site-effector interactions, allosteric signalling to one or more target functional sites, and optimisation of site-effector pairs by adjusting or tuning of site-effector composition and interactions, the method comprising the following steps:
 - viii. obtaining, for a protein of interest, 3D structure information and information on a protein function to be a target of allosteric modulation;
 - ix. identifying residues, of the protein of interest, that modulate one or more target functional sites of the protein of interest, by perturbing every residue or three-residue segment and quantifying allosteric signalling from each residue or three-residue segment to all other residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
 - x. identifying residues that modulate one or more target functional sites, by perturbing a functional site and quantifying allosteric signalling to all residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
 - xi. identifying a set of binding patches associated with residues that are allosterically linked to at least one of the one or more target functional sites as identified in steps ii and iii, and selecting one or more binding patches of interest from the set of binding patches by quantifying allosteric signalling from each binding patch to a target functional site, geometrical and chemical properties of the respective patch;
 - xii. for each binding patch of interest, identifying one or more ligands and corresponding binding sites;
 - xiii. either: ,

determining a set of allosteric site-effector pairs, each allosteric site-effector pair comprising a said ligand and a corresponding binding site as identified in step v, based on allosteric signalling from each pair to a said target functional site, the binding free energy, and physicochemical properties of the pair; or

if the one or more ligands do not possess a desired allosteric signalling and binding free energy, selecting a further one or more binding patches of interest and repeating from step v, and/or adjusting the compounds in step v, and repeating from step vi.

- xiv. outputting the set of site-effector pairs.
2. The method according to claim 1, wherein identifying one or more compounds comprises screening or designing a plurality of compounds for the binding patches, wherein the one or more compounds are selected from the plurality of compounds.
 3. The method according to claim 2, wherein step vi comprises ranking compounds in the plurality of compounds based on allosteric modulation and predicted binding free energy, and identifying the one or more compounds based on the ranking.
 4. The method according to claim 1, wherein identifying the one or more compounds comprises identifying one or more compounds that possess a desired allosteric effect and binding free energy.
 5. The method according to claim 1, wherein obtaining the 3D structure information comprises performing one or more of X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryogenic electron microscopy, or performing computational structural prediction.
 6. The method according to claim 1, wherein step ii involves simulating binding of a probe, using Structure-Based Statistical Mechanical Model of Allostery (SBSMMA), the model accounting for causality of allosteric communication and quantifying energetics of allosteric signalling and modulation, to each

three-residue segment in a protein chain of the protein of interest, and/or, separately, simulating mutations at all residue positions, and calculating the allosteric modulation to all residues.

7. The method according to claim 1, wherein step iii involves Reverse Perturbation (RP) to identify allosterically-linked protein residues by simulated binding at one of the target functional sites.
8. The method according to claim 3, wherein the processes of screening or designing compounds for binding patches and of ranking compounds based on allosteric modulation value and predicted binding affinity are repeated for two or more rounds.
9. The method according to claim 8, wherein, during each said round except for a final round, an analogue search and optimization process of obtained compounds is performed and an output of the process serves as inputs for a subsequent round.
10. The method according to claim 1, wherein steps ii, iii, iv, and vi involve use of Structure-Based Statistical Mechanical Model of Allostery (SBSMMA), the model accounting for causality of allosteric communication and quantifying energetics of allosteric signalling and modulation.
11. The method according to claim 10, wherein quantifying energetics of allosteric signalling and modulation involves an analysis of protein dynamics, an evaluation of per-residue allosteric potential, and a computation of the per-residue free energy associated with allosteric signalling based on a protein conformational ensemble of the protein of interest.
12. The method according to claim 11, wherein step vi comprises ranking the one or more site-effector pairs based on allosteric modulation, predicted binding free energy to the binding sites and physicochemical properties of site-effector pairs.

13. The method according to claim 12, wherein step vi comprises searching corresponding analogues and derivatives of the one or more compounds based on substructure matching or similarity and the ranking.
14. The method according to claim 8, wherein step iv comprises evaluating and prioritising the one or more binding patches based on allosteric modulation and a composition of residues.
15. The method according to claim 14, wherein the one or more binding patches are evaluated and prioritised based on properties of known allosteric sites in proteins and/or properties of known allosteric compounds for proteins.
16. The method according to claim 14, wherein the one or more compounds are evaluated and prioritised based on molecular weight, shape and aromaticity.
17. The method according to claim 14, wherein the one or more compounds are evaluated and prioritised based on properties of known allosteric sites in proteins and/or properties of known allosteric compounds for proteins.
18. The method according to claim 1, wherein for at least one of the one or more compounds, step vi comprises conducting a search of a chemical space surrounding the at least one compound, to identify further candidates of allosteric effectors based on substructure matching or similarity.
19. The method according to claim 1, wherein the one or more allosteric site-effector pairs are evaluated, prioritised and adjusted based on energetics of allosteric signalling and modulation at the one or more target functional sites, binding affinity and/or properties of site-effector compositions and interactions.
20. The method according to claim 19, wherein a simultaneous and mutual design of effector-site pairs with desirable properties comprises mutual adjustment of binding sites and effectors, via fine-tuning of the composition of allosteric site-effector pairs and intermolecular interactions of site-effector pairs, in respective adjustment, evaluation and prioritization steps across successive rounds.

21. The method according to claim 20, wherein evaluation, prioritisation and simultaneous and mutual adjustment of a site-effector pair continues in iterations until a site-effector pair with desired allosteric modulation and binding affinity is obtained.

22. A system for identifying at least one of allosteric drug candidates, their binding sites, site-effector interactions, allosteric signalling to one or more target functional sites, and optimisation of site-effector pairs by adjusting or tuning of site-effector composition and interactions, the system comprising:

a processor;

a memory device, accessible to the processor, the memory device comprising program code executable by the processor to:

- viii. obtain, for a protein of interest, 3D structure information and information on a protein function to be a target of allosteric modulation;
- ix. identify residues, of the protein of interest, that modulate one or more target functional sites of the protein of interest, by perturbing every residue or three-residue segment and quantifying allosteric signalling from each residue or three-residue segment to all other residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
- x. identify residues that modulate one or more target functional sites, by perturbing a functional site and quantifying allosteric signalling to all other residues, at single-residue resolution in the protein of interest, based on the 3D structure information;
- xi. identify a set of binding patches associated with residues that are allosterically linked to at least one of the one or more target functional sites as identified in steps ii and iii, and select one or more binding patches of interest from the set of binding patches by quantifying

allosteric signalling from each binding patch to a target functional site, geometrical and chemical properties of the respective patch;

- xii. for each binding patch of interest, identify one or more ligands and corresponding binding sites;
- xiii. either:
 - determine a set of allosteric site-effector pairs, each allosteric site-effector pair comprising a said ligand and a corresponding binding site as identified in step v,, based on allosteric signalling from each pair to a said target functional site, the binding free energy, and physicochemical properties of the pair; or
 - if the one or more ligands do not possess a desired allosteric signalling and binding free energy, select a further one or more binding patches of interest and repeating from step v, and/or adjusting the compounds in step v, and repeat from step vi.
- xiv. output the set of site-effector pairs.