



(51) International Patent Classification:
G01N 33/68 (2006.01) **C12Q 1/68** (2006.01)
G01N 33/15 (2006.01)

(74) Agent: **DAVIS, Bonnie, J.**; Taylor, Porter, Brooks & Phillips, P.O. Box 2471, Baton Rouge, LA 70821-2471 (US).

(21) International Application Number:
PCT/US2010/024777

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:
19 February 2010 (19.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/154,121 20 February 2009 (20.02.2009) US
61/261,939 17 November 2009 (17.11.2009) US

(71) Applicant (for all designated States except US): **BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE** [US/US]; 115 System Building, 3810 West Lakeshore Drive, Baton Rouge, LA 70808 (US).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **KIM, Sunyoung** [US/US]; 485 Clayton Court, Slidell, LA 70461 (US).

[Continued on next page]

(54) Title: PROTEIN STRUCTURAL BIOMARKERS TO GUIDE TARGETED CHEMOTHERAPIES

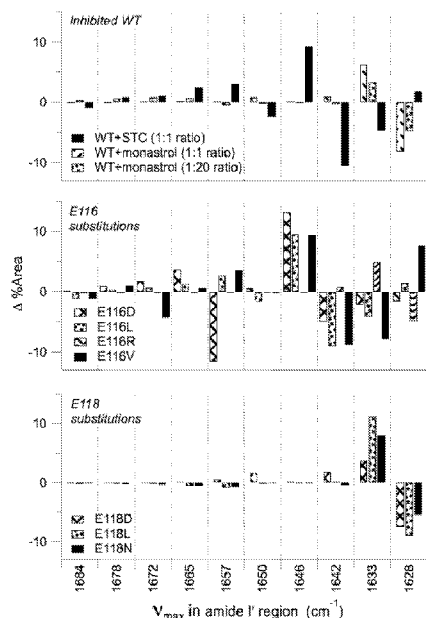


Fig. 4B

(57) Abstract: A rapid, infrared spectroscopic method has been developed to assess the efficacy of targeted chemotherapeutics against the structure of the polypeptide target, based on the effect of natural polymorphic sequence variation on the conformation of the protein. This method has an advantage over the current genomics-based screening, as the new method provides a direct readout of the structural, and hence functional, outcome of polymorphisms to the protein region targeted by drugs. It allows rapid measurement of a protein's susceptibility to therapeutic targeted agents, prior to using the drug as treatment in the patient. This method can be used to identify biomarkers for a response for a protein to a drug which can be readily tested, interpreted, and used in a clinical setting.

Published:

- *without international search report and to be republished
upon receipt of that report (Rule 48.2(g))*

**PROTEIN STRUCTURAL BIOMARKERS TO GUIDE
TARGETED CHEMOTHERAPIES**

Sunyoung Kim

File No: Kim 09M04W

[0001] The benefit of the filing dates of provisional U.S. application Serial Number 61/154,121, filed 20 February 2009, and of provisional U.S. application Serial Number 61/261,939, filed 17 November 2009, is claimed under 35 U.S.C. § 119(e) in the United States, and is claimed under applicable treaties and conventions in all countries.

[0002] The development of this invention was partially funded by the Government under grant number GM066328 from the National Institutes of Health. The United States Government has certain rights in this invention.

TECHNICAL FIELD

[0003] This invention pertains to a spectroscopic method to assess the efficacy of chemotherapeutics targeted to a specific protein in a patient, based on the effect of polymorphic sequence variation on the conformation pattern of the protein. This method views the structure of polypeptides and develops protein biomarkers that can be used to predict the protein's (and thus the patient's) response to a drug treatment. Using this method, protein biomarkers can be screened in individual patients to determine which patients will benefit from a particular drug therapy or to correctly tailor the dose of a drug and limit unnecessary toxicity.

BACKGROUND ART

[0004] Today's health practitioners are encouraged to diagnose and treat patients based on individual genetic composition. Pharmacogenomics promises to establish how genetic variation affects the way individuals will respond to drugs. Translation of this information into clinical settings would provide physicians the tools for selecting an optimal treatment for individual patients. Importantly, differences in drug responses caused by genetic differences among individuals could be correlated or quantified by diagnostic testing of an individual prior to receiving a particular drug or amount. The drug and amount could

thus be tailored to the individual. Biomarker tests are currently based on polymorphic alleles at the genetic and protein level. For example, Oncotype DX is a commercially available test that predicts breast cancer recurrence and is used as a treatment guide for chemotherapy for breast cancer.

[0005] Pharmacogenomics research focuses primarily on establishing correlations between microarray gene expression profiles and disease states, often without defining biochemical pathways and mechanisms. This strategy is widely adopted, easy to automate, and scalable to high-throughput modes. However, the reliability of microarray biomarker assays for disease is unknown with few effective biomarkers or tools for assessment established. For example, only one assessment has begun to be used in oncology practice.

[0006] Efforts to identify cancer biomarkers through genomic methods have included use of microarray assays to mine for genetic predictors of therapeutic response and to mine for variations in tumor gene expression in patients whose disease later exhibits different degrees of aggressive behavior. These biomarkers are defined not by classical single markers, but by sets of genes whose up- or down-regulation in expression is correlated with a therapeutic response or aggressive behavior of the disease. Many of the gene signatures generated in the last five years for predicting disease have had only marginal overlap with those identified in the original studies, even though subsequent work had corroborated pathway mechanisms in the model organism studies. For example, BCR-ABL kinase domain mutations have been found to confer resistance to the tyrosine kinase inhibitor imatinib, a competitive inhibitor for the substrate ATP (60,61,62). As used herein and in the claims, a “drug” or “drug compound” is a therapeutic compound that is given to a patient that targets a specific protein.

[0007] Minimizing problems of validity, reproducibility, and bias in biomarker generation and employment is imperative when searching for predictors of therapeutic response. There is a need for standardization in methods and technology platforms to define biomarkers for predicting disease progression as well as predicting response. The interaction of a targeted protein and its inhibitor drug can be useful. In particular, certain enzymes such as ATPases and kinases that use ATP as a substrate have showed potential as targets for drugs. ATPases are enzymes that use ATP hydrolysis to achieve a cellular function, and are involved in various human diseases. Kinases will transfer the phosphate group in ATP to another compound. As such, ATPases and kinases are important protein targets for

therapeutic drugs. In the ATPases, most of the drugs have inhibited activity by binding to an allosteric site. Examples include, Kinesin (Eg5) and its inhibitor monastrol used for cancer therapy; muscle myosin II and its inhibitor *N*-benzyl-*p*-toluene sulphonamide used for muscle relaxation; Na⁺,K⁺-ATPase and the inhibitor digoxin used for heart failure; P glycoprotein and its inhibitor XR9576 used for cancer, H⁺,K⁺- pump and its inhibitor benzimidazoles used for gastric ulcers; vacuolar H⁺-ATPase and the inhibitor bafilomycin A1 used for osteoporosis; and DNA topoisomerase II and the inhibitor ICRF-193 used for cancer (59).

[0008] Allostery is important in controlled catalysis, signal transduction, and apoptosis (1). The classic view of proteins demonstrating this property (2) asserts that binding of a ligand at one site provokes conformational changes at a remote, second site. Recent studies (3) evaluating underlying mechanisms of allostery alternatively suggest that ligand binding results in selection of pre-existing conformational substates. Implicit in the latter model, interactions between the orthosteric and allosteric sites are tightly linked through structure and thermodynamics (4).

[0009] The human Kinesin-5 motor protein (Eg5 or KSP) plays key roles in bipolar mitotic spindle formation and is a protein target for allosteric compounds (5-7) that alter catalytic ATPase activity of the protein (8,9). There are many compounds and classes of compounds that are known to inhibit KSP including monastrol, S-trityl-L-cysteine and its derivatives, quinazolines (e.g., ispinesib), adociasulfates, tetrahydroisquinolines, dihydropyrazoles, thiophenes, pyrrolotriazines, thiazoles, gossypol, indoles, and biphenyls. The best-characterized inhibitors, monastrol (10) and S-trityl-L-cysteine [STC; (9)], were uncovered from independent chemical screens. Biochemical studies demonstrate a wide concentration range of inhibition by these compounds (10-12) differences in the kinetic mechanism of allostery (13-15), and even allosteric activation (16) is possible.

[0010] Interest in these allosteric compounds, monastrol and STC, has been acute as they are potential anticancer agents. Additionally, these compounds serve as research tools to probe the fundamental mechanism by which Eg5, and perhaps all other motor proteins, convert and transduce energy to conformational changes in distal regions of the protein. Insights concerning allosteric conformational states of Eg5 result principally from diffraction- and microscopy-based techniques. The kinesin motor domain is an arrowhead-shaped structure with a central β -sheet flanked by 3 helices on each side. In crystallographic studies of Eg5•ADP complexed with monastrol (17,18) and other allosteric inhibitors (12,19-21), the

wildtype Kinesin-5 motor domain displays a similar conformer, irrespective of the chemical nature of the allosteric drug.

[0011] The most notable conformational change observed is the adoption of a 'closed' conformation by the insertion loop (L5) of $\alpha 2$ helix that cradles the allosteric compound. This is in contrast with the 'open' conformation observed in the absence (22) of an allosteric ligand. However, conformational transitions of the L5 loop are found not only in response to drug binding, but also in normal motor function. The Eg5 conformer trapped by allosteric agents is suggested to be an intermediate state of its normal ATP hydrolysis cycle (23). The L5 loop is observed in the 'closed' conformation in cryo-EM experiments (24) with a *D. melanogaster* homologue of Eg5 bound to microtubules using non-hydrolyzable substrate analogues and in Eg5•AMPPNP crystals in the absence of allosteric agents (25).

[0012] No consensus of L5 residues key for transmitting allosteric information has been established, primarily due to interdependence of the allosteric and active-site interactions and the lack of rapid methods to observe structures of mutant motor proteins. Simple sequence conservation analysis does not reveal how the L5 loop specifically recognizes and binds an allosteric molecule. Studies (15,17,26,27) using targeted mutagenesis to measure the contribution of specific contacts in Eg5 allostery and chemical-kinetic measurements of mutant motor ensembles in solution on the whole concluded that their data were consistent with the crystallographic observation of the wildtype Eg5•inhibitor complexes. Yet, neither these reports nor X-ray structural analyses of allosterically-inhibited Eg5 proteins arrive at a common set of contact residues for binding many different chemical partners.

[0013] The advent of the structural genomics era has led to the curation of many motor protein structures; to date, 20 RCSB Protein Data Bank (PDB) entries for wildtype Eg5 alone have been deposited. (See the website: <http://www.rcsb.org/pdb>) Although the atomic level detail of the structures is fairly complete, whether these static 'snapshots' accurately reflect functional states that these amino acid assemblies can achieve is more difficult to determine. Thus, obtaining all needed conformational possibilities in kinesin mechanotransduction and its allosteric inhibition is a significant challenge.

DISCLOSURE OF INVENTION

[0014] I have developed an infrared spectroscopic method to assess the efficacy of targeted chemotherapeutics against the structure of the polypeptide target based on the effect of natural polymorphic sequence variation on the conformation of the protein. This method has an advantage over the current genomics-based screening, as the new method provides a direct readout of the structural, and hence functional, outcome of polymorphisms to the protein region targeted by drugs. Second, it allows rapid measurement of tumor protein susceptibility to therapeutic targeted agents, prior to using the drug as treatment in the patient. Third, this method can be used to identify biomarkers for a response for a protein to a drug which could be readily tested, interpreted, and used in a clinical setting.

[0015] For an example of the above discovery, the human Kinesin-5 protein (Eg5 or KSP), essential in mitosis, was used. Eg5 is known to be a target for >80 classes of allosteric compounds that bind to a surface-exposed site formed by the L5 loop. Ligand-bound states of two L5-directed inhibitors against wildtype and 15 Kinesin-5 mutants were compared by ATPase assays and Infrared (IR) spectroscopy. Biochemical kinetics uncovered functional differences between individual residues at the N- or C-termini of the L5 loop. Infrared evaluation of solution structure and multivariate analysis of the vibrational spectra uncovered changes in L5-localized 3_{10} -helix and disordered content, regardless of substitution or drug potency. Coupled to these local structural events are two types of rearrangements in β -sheet hydrogen-bonding. Drug potency was correlated with these transformations in β -sheet contacts and was corroborated by wildtype Kinesin-5 crystal structures. These data directly show that mutation and/or ligand binding can not only remodel the allosteric binding surface but also transmit long-range effects. These findings prove that these relatively rapid IR approaches can find and identify structural biomarkers that can be used in a determination of drug sensitivity and drug efficacy.

[0016] The invention provides a method for predicting the probable response of a patient to a drug compound; wherein the patient has a disease; wherein the drug compound inhibits a protein that is associated with the disease; and wherein the protein is polymorphic in the population, such that there is substantial variation in the responsiveness of the disease to the drug compound between individuals as a result of such polymorphism; said method comprising the steps of: **(a)** determining whether the variant of the protein expressed by the patient has greater spectroscopic similarity to: **(i)** a variant of the protein that is known to be

responsive to the drug compound, in the absence of the drug compound, or (ii) a complex of the known-responsive-variant of the protein bound to the drug compound; and (b) predicting that the patient will likely be responsive to the drug compound if there is greater spectroscopic similarity to (i); or predicting that the patient will likely be unresponsive or less responsive to the drug compound if there is greater spectroscopic similarity to (ii); wherein the degree of spectroscopic similarity is determined as infrared absorption or transmittance at one or more wavelengths at which there is a significant difference in absorption or transmittance between (i) and (ii).

[0017] In the method given above, the determination of the variant of a protein expressed by the patient can be conducted by direct spectroscopic measurement of protein expressed by the patient. In particular, such determination can be made by genetic analysis of the patient's DNA or RNA encoding the protein, and comparing the patient's allele or alleles for the protein to those of one or more reference samples for which direct spectroscopic measurements of the protein have previously been conducted. Any spectroscopic method that measures molecular vibrations at different frequencies can be used, e.g., infrared spectroscopy or Raman spectroscopy. A preferred method is Fourier-transform infrared spectroscopy is useful.

[0018] This method can be used for patients with cancer or other diseases, and with drugs targeted to the specific disease, such as chemotherapeutic agents. One such chemotherapeutic agent would be one that acts as an allosteric inhibitor of the protein. One such protein that can be targeted is the kinesin Eg5 protein. Two known drugs that inhibit mitosis by inhibition of the Eg5 protein are monastrol and S-trityl-L-cysteine (STC).

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] Fig. 1A illustrates the results of the wildtype Eg5 and Eg5 proteins with substitutions at the E116 residue, as described in Table 1, as analyzed by SDS-PAGE using a 10% acrylamide gel. The gel migration of the double-mutant, E116D/E118D (DD), matched wildtype Eg5 samples (data not shown).

[0020] Fig. 1B illustrates the results of the wildtype Eg5 and Eg5 proteins with substitutions at the E118 residue, as described in Table 1, as analyzed by SDS-PAGE using a 10% acrylamide gel.

[0021] Fig. 1C illustrates the results of the wildtype Eg5 and Eg5 proteins with substitutions at the D130 residue, as described in Table 1, as analyzed by SDS-PAGE using a 10% acrylamide gel.

[0022] Fig. 1D illustrates the results of the wildtype Eg5 and Eg5 proteins with substitutions at the A133 residue, as described in Table 1, as analyzed by SDS-PAGE using a 10% acrylamide gel.

[0023] Fig. 2A illustrates the normalized ATPase activity expressed as a percent of wildtype for the mutant proteins for E116 and E118 as described in Table 1. Shown are steady-state, basal ATP hydrolysis rates, for which mean $V_{\max}$ for WT (black) is $0.144 \pm 0.003 \text{ s}^{-1}$. For WT, $n = 101$, while n for all substitutions ranges from 20-51. DD represents the double mutant E116D-E118D. All reactions contain 3-15 mM NaCl.

[0024] Fig. 2B illustrates the normalized ATPase activity expressed as a percent of wildtype for the mutant proteins for D130 and A133 as described in Table 1. Shown are steady-state, basal ATP hydrolysis rates, for which mean $V_{\max}$ for WT (black) is $0.144 \pm 0.003 \text{ s}^{-1}$. For WT, $n = 101$, while n for all substitutions ranges from 20-51. All reactions contain 3-15 mM NaCl.

[0025] Fig. 2C illustrates a plot of normalized Eg5 ATP hydrolysis rates for E116 mutants against increasing concentrations of monastrol (left panels) or STC (right panels). WT kinetic data (black or grey) are superimposed on E116. ATP hydrolysis rates for each substitution are normalized with respect to the 0 mM monastrol or 0 mM STC measurements of the parent substitution. All data points in the figure reflect the averaged normalized rate of at least 3 independent trials and 2-4 separate protein preps and the standard error.

[0026] Fig. 2D illustrates a plot of normalized Eg5 ATP hydrolysis rates for D130 mutants against increasing concentrations of monastrol (left panels) or STC (right panels). WT kinetic data (black or grey) are superimposed on D130. ATP hydrolysis rates for each substitution are normalized with respect to the 0 mM monastrol or 0 mM STC measurements of the parent substitution. All data points in the figure reflect the averaged normalized rate of at least 3 independent trials and 2-4 separate protein preps and the standard error.

[0027] Fig. 2E illustrates a plot of normalized Eg5 ATP hydrolysis rates for E118 mutants against increasing concentrations of monastrol (left panels) or STC (right panels).

WT kinetic data (black or grey) are superimposed on E118. ATP hydrolysis rates for each substitution are normalized with respect to the 0 mM monastrol or 0 mM STC measurements of the parent substitution. All data points in the figure reflect the averaged normalized rate of at least 3 independent trials and 2-4 separate protein preps and the standard error.

[0028] Fig. 2F illustrates a plot of normalized Eg5 ATP hydrolysis rates for A133 mutants against increasing concentrations of monastrol (left panels) or STC (right panels). WT kinetic data (black or grey) are superimposed on A133. ATP hydrolysis rates for each substitution are normalized with respect to the 0 mM monastrol or 0 mM STC measurements of the parent substitution. All data points in the figure reflect the averaged normalized rate of at least 3 independent trials and 2-4 separate protein preps and the standard error.

[0029] Fig. 3A illustrates the method of band narrowing and regression analysis for wildtype 2nd derivative (solid), Fourier deconvolution (dashed), and peak fit calculations. The amide I' band of the averaged FTIR spectra (solid) is superimposed with the fitted trace (dashed light grey). Areas under the Lorentzian components were compiled and wildtype subtracted to result in spectrotypes profiles of the structural data. Frequencies of the spectral components for all spectra were centered and binned to ± 2 wavenumbers.

[0030] Fig. 3B illustrates the method of band narrowing and regression analysis for E116L mutant 2nd derivative (solid), Fourier deconvolution (dashed), and peak fit calculations. The amide I' band of the averaged FTIR spectra (solid) is superimposed with the fitted trace (dashed light grey). Areas under the Lorentzian components were compiled and wildtype subtracted to result in spectrotypes profiles of the structural data. E116L showed a loss at 1684 and a gain at 1644. Frequencies of the spectral components for all spectra were centered and binned to ± 2 wavenumbers.

[0031] Fig. 3C illustrates the method of band narrowing and regression analysis for E116R mutant 2nd derivative (solid), Fourier deconvolution (dashed), and peak fit calculations. The amide I' band of the averaged FTIR spectra (solid) is superimposed with the fitted trace (dashed light grey). Areas under the Lorentzian components were compiled and wildtype subtracted to result in spectrotypes profiles of the structural data. Band narrowing of E116R resulted in Lorentzians centered at the same frequencies as WT. Both do not contain a 1646 Lorentzian component. Frequencies of the spectral components for all spectra were centered and binned to ± 2 wavenumbers.

[0032] Fig. 4A illustrates plots of the averaged Fourier-transform infrared (FTIR) spectra of the amide I' region of wildtype, E116 mutant, and E118 mutant polymorphisms. The wildtype spectra are superimposed (dotted line) on each trace and differences in line shape are highlighted, showing quantitative changes in secondary structure. FTIR spectra were acquired from two independent purifications and averaged.

[0033] Fig. 4B illustrates the results of band narrowing of the amide I' region of the N-terminal L5 residues into 10 frequency bins shown on the horizontal axis. Positive amplitudes indicate a gain of secondary structure at a given frequency from wildtype, while negative amplitudes indicate a loss of structure. Top panel shows wildtype as inhibited by both monastrol in 1:1 (black box, dark grey fill) and 1:20 (black box, mid grey fill) ratios and STC (light grey box, dot fill). Middle panel shows mutant E116 polymorphisms as compared to wildtype. Bottom panel shows mutant E118 polymorphisms as compared to wildtype.

[0034] Fig. 5A illustrates plots of the averaged Fourier-transform infrared (FTIR) spectra of the amide I' region of wildtype, D130 mutant, and A133 mutant polymorphisms. The wildtype spectra are superimposed (dotted line) on each trace and differences in line shape are highlighted, showing quantitative changes in secondary structure. FTIR spectra were acquired from two independent purifications and averaged.

[0035] Fig. 5B illustrates the results of band narrowing of the amide I' region of the N-terminal L5 residues into 10 frequency bins shown on the horizontal axis. Positive amplitudes indicate a gain of secondary structure at a given frequency from wildtype, while negative amplitudes indicate a loss of structure. Top panel shows wildtype as inhibited by both monastrol in 1:1 and 1:20 ratios and STC (light grey box, dot fill). Middle panel shows mutant D130 polymorphisms as compared to wildtype. Bottom panel shows mutant A133 polymorphisms as compared to wildtype.

[0036] Fig. 6A illustrates the results of PCA analysis for the E116 and E118 mutant proteins, the N-Terminal L5 mutants. The score, or spatial positioning, of the data points is based on changes in percent area of the secondary structure profiles determined by band narrowing. PCA clustered the majority of E118 (■) and WT (●) data points based on secondary structure. The E116 (□) polymorphisms fell into a separate line of degeneracy with a larger disordered component. The 1642 and 1646 cm^{-1} vectors separate these two populations.

[0037] Fig. 6B illustrates the results of PCA analysis for the D130 and A133 mutant proteins, the C-Terminal L5 mutants. The score, or spatial positioning, of the data points is based on changes in percent area of the secondary structure profiles determined by band narrowing. PCA of C-terminal L5 residues, A133 ($\blacktriangle$) and D130 (Δ), fell predominantly into the same population occupied by the N-terminal E118 mutations and can be differentiated by the 1633 and 1628 cm^{-1} vectors, respectively.

[0038] Fig. 6C illustrates the results of PCA analysis for a composite of both N- and C-terminal polymorphisms. Highlighted are the monastrol and STC data points.

MODES FOR CARRYING OUT THE INVENTION

[0039] I have shown that the efficacy of targeted chemotherapeutics against the structure of polypeptide targets can be directly assessed through a pioneering high-throughput, specialized implementation of vibrational spectroscopy. The final cellular target for clinical drugs is typically polypeptides but correlations between protein structure and therapeutic outcome are commonly unknown. There are no reports of a rapid and quantitative method for assaying conformational alterations resulting from peptide sequence variations. This lack of information has hampered translation of gene signatures into clinical diagnostics. Here I have found protein spectroscopic signatures that reflect conformational ensembles adopted in solution and that can be used as clinical biomarkers that can improve treatment decisions.

[0040] Using this method, the efficacy of targeted chemotherapeutics for patients can be predicted, potentially even in a quantitative manner, based on the polymorphic variant of the targeted protein in the individual and based on the effect of this polymorphic sequence variation on protein conformation. This method is unique on three fronts. First, protein conformation has an advantage over genomics-only-based evidence, as it provides a direct readout of the structural, and hence functional, outcome of polymorphisms to the actual target protein for drugs. Second, this method allows rapid measurement of a target protein's susceptibility to therapeutic agents, rather than deciphering treatment effectiveness from long-term observation of disease progression. Third, a single biomarker for response of a protein target to a drug, independent of tumor state or stage, is readily testable, interpretable, and amenable to translating into clinical application, compared to complex, multivariable signatures of gene or protein networks.

[0041] Clinicians, for the first time, will be able to reliably predict which individual patients will benefit from a particular drug treatment, to correctly tailor dosages, and to limit unnecessary toxicity, based on rapid assessment of drug-to-target interactions. Most cancer treatments at present benefit only a minority of patients to whom they are administered, and this over-treatment of patients is a major expense for society. A second aspect of this work is that structural screening of target proteins may be extendable to established linkages between polymorphisms and drug efficacy for patients with neurodegenerative or heart diseases.

[0042] One embodiment of this invention is to focus on treatment to traditional cytotoxic drugs that affect rapidly dividing cells in general, and uses small-molecule inhibitors specific for the deregulated proteins of cancer cells. These deregulated proteins are often enzymatic domains that are typically mutated and/or overexpressed specifically in cancer cells. Oncological protein targets catalyzing nucleotide triphosphate hydrolysis, such as kinesin motor proteins and tyrosine kinases, are examples of such proteins. These protein families share a common catalytic capacity of hydrolyzing adenosine triphosphate (ATP).

[0043] Protein kinases use one of the products of ATP hydrolysis to phosphorylate protein substrates, and thereby trigger signal transduction cascades in cells. These enzymes regulate multiple processes that contribute to tumor development and progression, including cell growth, differentiation, migration, and apoptosis. For human tumors whose growth is driven by these activated kinases, targeted drugs, such as imatinib (Gleevec), can inhibit or reverse malignant progression. In contrast, kinesin motor proteins utilize the energy from ATP breakdown to move along or destabilize microtubules. Essential for cell division, the clinical and oncogenic role of kinesin spindle protein (KSP) in mitosis has been demonstrated by a number of groups: (i) inhibition of KSP leads to apoptotic cell death and (ii) KSP is preferentially over-expressed in malignant cells, but not normal non-dividing cells such as terminally differentiated neurons. From the >80 chemical classes of KSP inhibitors reported, five compounds, including ispinesib, are under active clinical development by pharmaceutical companies.

[0044] I have developed a high-throughput method toward analyzing protein structure using FTIR in solution. In this post-genomic era, hundreds of protein variants can be produced cheaply and quickly. It has been much less straightforward to screen or select for protein structure: X-ray crystallography, NMR spectroscopy, and even CD spectroscopy are not amenable to high-throughput approaches. The protein secondary structure can be

monitored by Fourier-transform infrared (FTIR) spectroscopy. The amide I', or C=O stretching, IR vibration of the peptide bond is sensitive to the hydrogen bonding patterns of differing secondary structure elements and depends on the solution structure of proteins.

[0045] I have developed a high-throughput measurement of solution structure with the motor domain of the human kinesin spindle (KSP) protein. The motor domain of the human KSP is the 'business end' of the kinesin molecule: it contains the binding sites for ATP and microtubules. It also provides several experimental advantages when compared to other proteins: (i) it is a relatively small protein that can be produced in large quantities (15-60 mg KSP/L culture) using recombinant expression systems; (ii) basal ATPase activity is readily measured in high-throughput format; and (iii) such systems are amenable to efficient site-directed mutagenesis and crystallization.

[0046] As shown below, IR measurements were used to discover a structural biomarker that can be used for biophysical profiling of polymorphisms at the protein level. More than half of the mutants examined had catalytic rates comparable or greater than wildtype samples. I predict that patients with seemingly normal Eg5 function can harbor a polymorphism that would confer resistance to clinical drugs. The method described herein using the structural biomarker screen can help an oncologist find the best inhibitor for the Eg5 of a specific patient.

[0047] This invention uses vibrational spectroscopy to monitor protein structure biomarkers to predict drug efficacy for targeted chemotherapeutic agents. Although these biophysical methods are low resolution, in comparison to x-ray crystallography or NMR spectroscopy, the experimental advantage is its realization in high-throughput modes. There is no parallel approach to microarray genetic profiling that can probe protein structure. Results from conformational profiling will allow the scientific community to evaluate subtle as well as pleiotropic effects to protein structure upon amino acid substitution and build a large database of conformational states to query additional scientific questions. For example, this method could assist in optimizing patient selection for clinical trials: individuals who were found to be resistance to the test drug in a prescreen could be eliminated from the trial and data obtained on remaining patients would be a truer representation of the trial drug efficacy. In addition, such biophysical profiling of kinesin conformational ensembles, shown below, may be extendable to the other kinesin proteins, such as Kif5A, Kif21A, and Kif1B, in

which polymorphisms have been demonstrated to result in neurodegenerative conditions, and Kif6 in acute coronary syndromes. (55,56,57,58)

[0048] One embodiment of this new method is on chemotherapeutic inhibitors such as allosteric inhibitors that target specific proteins of cancer cells, e.g., the use of monastrol on human kinesin Eg5 (also known as KSP or kinesin spindle protein). The first step is to assay *in vitro* conformational signatures of the purified protein by vibrational spectroscopy, e.g., by using Fourier-transform infrared (FTIR) spectroscopy; comparison of the 1700-1600 cm^{-1} region of the infrared spectrum from wildtype protein alone and from wildtype protein bound to an inhibitor will detect changes in wild-type protein secondary or tertiary structure due to the binding of the inhibitor. Then using the same technique, a protein with a known polymorphism can be assayed for secondary or tertiary structure changes. I have found that variant proteins that are resistant to the inhibition of the drug often show secondary or tertiary structural changes similar to those seen in wildtype proteins when bound to the chemotherapeutic drug, e.g., as revealed by alterations in the IR spectrum. The variant proteins were resistant to inhibition because their secondary or tertiary structure was already similar to that of the “inhibited” state, even though the proteins may not be functionally inhibited. Tumors with these variant proteins would not be affected by treatment with the chemotherapeutic drug. Thus using vibrational spectroscopy, the efficacy of treating a certain tumor in a patient with a chemotherapeutic drug that targets a specific tumor protein can be assayed quickly prior to treatment.

[0049] One example is the differing efficacies in drug inhibition of Eg5. Without wishing to be bound by this theory, I believe that there is more than one pathway for allosteric communication that initiates from the L5 loop of Eg5: differences in the inhibitory power of monastrol and STC amongst the Kinesin-5 orthologues thus result from these deviations in allosteric communication. Biochemical behavior of mutant proteins has been measured as well as the protein secondary structure in solution. The resulting model reveals that allosteric conformational changes at the protein surface can promote two different types of intraprotein contacts at the central core of the kinesin motor domain. These changes discriminate between the inhibitory power of these allosteric drugs.

Example 1

Materials and Methods

[0050] *Generation of single- and double-mutants of Eg5.* Starting cDNA was the truncated form of wildtype Eg5 as previously reported [residues 1-370; (28)]. Appropriate primers were synthesized and used for site-directed mutagenesis with the QuikChange II XL kit (Stratagene, La Jolla, California). Construction of the Eg5 motor domain with substitutions of Asp for E116 and E118 was performed using two rounds of mutagenesis. Expected mutations were confirmed by a single-read sequencing reaction. A subset of five mutations was completely sequenced on both strands to ensure that inadvertent coding errors were not introduced. The mutations produced are shown below in Table 1.

[0051] *Motor protein expression and purification.* The wildtype Eg5 kinesin motor domain, as well as single- and double-site polymorphisms, were expressed in BL21-Codon Plus (DE3)-RIL cell lines (Stratagene) and purified by cation exchange chromatography as previously described (28). Purity of wildtype and all Eg5 substitutions were determined by SDS-PAGE analysis using a 10% acrylamide-0.27% bisacrylamide gel, with a 0.38 M Tris-HCl, pH 8.8 resolving and a 0.125 M Tris-HCl, pH 6.8 stacking gel components. SDS sample buffer was added, adjusting the sample to a final concentration of 2% SDS, 0.29 M 2-mercaptoethanol, 0.05 M Tris-HCl pH 6.8, 10% glycerol, 0.0025% (w/v) bromophenol blue. SDS-PAGE samples were then boiled at 100°C to ensure denaturing conditions. Following boiling, samples were immediately run along an unstained molecular weight ladder (New England Biolabs, Massachusetts) and were visualized using Sypro Tangerine (Molecular Probes, Oregon). Protein purity was estimated to be >90%. All protein concentrations were determined by Bradford assay (Coomassie Plus Protein Assay Reagent, ThermoSScientific, Massachusetts), using bovine serum albumin (BSA) as the standard; concentrations ranged from 0.5–6.0 mg/mL. Final storage buffer for the purified Eg5 proteins was 125 mM NaCl, 50 mM HEPES pH 7.4, 0.1 mM MgATP, 1 mM DTT and 10% glycerol. Purified motor protein was flash-frozen in liquid nitrogen and stored at -80°C until use. Unless otherwise indicated, all chemical compounds were purchased from Sigma-Aldrich (St. Louis, Missouri).

[0052] *In vitro ATPase assays.* Basal ATPase activities were measured with a coupled enzymatic assay as previously described (29) with 5 μ M motor protein and 1 mM MgATP in a SpectraMax 2E spectrophotometer for high-throughput kinetic measurements as previously described (30). Reactions were supplemented with monastrol (0–600 μ M; Sigma-

Aldrich, St. Louis, Missouri) or S-trityl-L-cysteine (STC; 0–200 μ M; Sigma-Aldrich, St. Louis, Missouri) as appropriate, and control reactions contained equivalent dimethyl sulfoxide (DMSO). Reactions used for calculation of the IC_{50} values contained between 5 and 22.5 mM NaCl. Data were reproducible with protein samples purified from differing batches of cultured cells from two different laboratories.

[0053] *Infrared spectroscopy.* All spectroscopic data were acquired at room temperature as previously described (28) on a Vertex 80v (Bruker Optics, Billerica, Massachusetts) with the use of a BioATR (Bruker Optics, Billerica, Massachusetts). The spectrometer was equipped with a liquid nitrogen-cooled MCT (mercury cadmium telluride) cryodetector or photovoltaic detector and a KBr beam splitter. The mirror velocity was 20 kHz, the spectral resolution was 2 cm^{-1} , and 1000 mirror scans were acquired for each double-sided interferogram. A 6 mm aperture, gain of 1 on the preamp A setting, Blackman-Harris 3-Term apodization function, and two levels of zero-filling were used. A minimum of 3 spectra and two independent protein purifications per Eg5 sample were used to generate the averaged data.

[0054] Protein samples were 2H_2O -exchanged using Zeba desalting spin columns (Pierce, Rockford, Illinois) or SpectraPor dialysis membranes (Spectrum Laboratories, California) and equilibrated with 2H_2O buffer (10 mM HEPES p²H 7.4, 0.5 mM $MgCl_2$, and 0.1 mM $MgATP$). Protein samples utilized in IR experiments had final concentrations of 0.9–5 mg/ml, as determined by Bradford assays. Stock concentrations of monastrol and STC were 100 mM in 100% DMSO. Note that positive displacement pipettes were used for all DMSO aliquotting to ensure accurate volumetric measurement of the viscous solvent. For Eg5•inhibitor samples, a 20 μ l sample was achieved by mixture of 19 μ l of Eg5 protein with 1 μ l of inhibitor in equimolar concentrations, resulting in a final DMSO concentration of 5%. Inhibitor stocks were serially diluted in DMSO, using a maximum 1:5 ratio, until the desired concentration was achieved. Of this 20 μ l mixture, a sample volume of 12–15 μ l was dispensed onto the ATR crystal and allowed to equilibrate to room temperature for 10–15 minutes before an interferogram was collected.

[0055] *Spectral analyses.* Each sample spectrum was subtracted from the spectrum of 2H_2O buffer or 2H_2O buffer plus DMSO when samples contained drug. After baseline correction, spectra were area normalized. Savitsky-Golay second-derivative, Fourier self-deconvolution, and spectral fitting analysis were performed with Grams/AI (ThermoFischer

Scientific, Massachusetts). Self-deconvolution was performed with a Bessel apodization function and peak fitting utilized the Levenberg-Marquardt method with Lorentzian components. Identical parameters were used for all spectra with an amide I' fitting range of 1710 cm^{-1} to 1594 cm^{-1} . Bandwidth, frequency, and amplitude of each spectral component were iterated until a minimum in the fit to the experimental data was obtained. The frequencies of the resulting spectral components and their amplitudes generated good fits, as judged by residuals and reduced χ^2 parameters.

[0056] The frequencies, obtained from the second-derivative and Fourier self-deconvolution analyses, were used as the first set of inputs for the multivariate data analyses. Area-normalized data sets of infrared spectra, in comparison to wildtype Eg5 samples, also were used as the second set of inputs for the multivariate data analyses. Principal component analysis (PCA) was performed with Matlab software (Mathworks, Massachusetts). PCA plots presented were scaled on x- and y-axis [0,1]. The point spread on the PCA indicated that the infrared measurements sample the variances in vector space, and there is high signal-to-noise in the infrared spectra and calculations. All figures were generated using IGOR Pro software (Wavefunction, Oregon).

[0057] *Crystallization and X-ray structure solution.* The Eg5 motor domain (residues 1-369) was expressed in bacteria and purified as previously described (25). Prior to crystallization experiments, purified Eg5 (~ 250 μM) was mixed with 1 mM STC (Sigma Aldrich). Original crystals obtained in 250 mM NH_4SO_4 , 25% PEG 3350, 100 mM MES pH 6.0, and 10 mM trimethylamine hydrochloride were used to streak seed crystals grown in 250 mM NH_4SO_4 , 25% PEG 3350, 100 mM MES pH 6.0, 10 mM trimethylamine hydrochloride, and 10-15% glycerol. Crystals were grown at 4°C in sitting drops consisting of 2 μl of protein/drug mixture and 1 μl of well solution. Cubic crystals appeared several days after streak seeding. The crystals were allowed to grow for several weeks before being flash frozen for data collection. Diffraction data were collected at 100K using a Bruker Microstar X-ray generator equipped with Helios optics and a Proteum 4K Platinum 135 CCD camera and were integrated and scaled with PROTEUM2 software (Bruker AXS). Crystals belonged to space group I2(1)3 with cell parameters of $a=b=c=157.93\text{\AA}$. The Eg5•STC complex structure was determined via molecular replacement methods (AMoRe) using a search model based on molecule A of the Eg5•ADP•monastrol complex (PDB ID 1X88). Successive rounds of manual model building as previously described (31) and automated refinement

were performed using O and CNS. The final model was refined to R and R_{free} values of 24.07 and 27.46, respectively (All $|F| > 0$; 25.0-2.5 Å), and contained residues 18-271 and 288-366, along with the Mg-ADP complex, 74 waters, one STC molecule, and one trimethylamine molecule.

Example 2

Substitutions of L5 residues result in altered SDS-PAGE mobility

[0058] The initial step of this study was to assess how perturbations of the L5 loop affect Eg5 steady-state kinetics and solution structure. The first type of perturbation involves alteration of sidechain chemistry. Fifteen (15) mutations were sampled at four sites in the L5 loop: two at the N-terminus and two at the C-terminus of the insertion loop. Prior studies scrutinizing the allosteric pocket typically assumed that point mutations have little effect beyond the localized change of a particular functional group. However, this stance can underestimate the ability of the point mutation to remodel the binding surface or transmit long-range effects that are dependent on the nature of the substitution. Therefore, the mutations used herein ranged from conservative to pleiotropic alterations; this allowed a wider range of chemical and structural outcomes at specific positions to be seen.

[0059] The N-terminal residues chosen were E116 and E118. The E116 sidechain is oriented inwards toward the allosteric pocket, in contrast to the E118 sidechain orientation toward the orthosteric site. C-terminal L5 residues studied were D130 and A133, embedded within short stretches of negatively-charged and hydrophobic sidechains and interrupted by a proline. There was no apparent interaction between the sidechains of any of these four residues and the allosteric inhibitors in co-crystal structures. However, the E116V and E118N substitutions are present in the *Drosophila* Kinesin-5 protein that is insensitive to allosteric inhibitors and that the D130V and A133D mutations were found in human cell lines resistant to Eg5 allosteric inhibitors (16).

[0060] Expression plasmids with these and other substitutions in Eg5 cDNA were successfully constructed, and the mutant motor proteins were readily expressed and purified, as evident from SDS-PAGE (Figs. 1A-1D). The location and amino acid substitutions for the mutant proteins in Figs. 1A-D are shown in Table 1. Wildtype Eg5 has an apparent and expected molecular weight of 42 kDa. Migration of the mutant Eg5 proteins in SDS-polyacrylamide gel electrophoresis, however, does not correlate with calculated molecular

weights. Both forms of anomalous gel mobility are observed for mutant proteins. Faster migration of purified proteins is detected for nonconservative substitutions of the L5 carboxylates (Figs. 1A and 1B). C-terminal mutant proteins overall have slower gel mobility in SDS-PAGE (Figs. 1C and 1D).

Table 1: Location and Amino Acid in L5 Loop of Human Eg5		
Name-Location in Eg5	Wildtype (WT) /Mutation (M)	Amino Acid
A. N terminal at 116 location		
E116	WT	Glutamic Acid
E116D	M	Aspartic Acid
E116L	M	Lysine
E116R	M	Arginine
E116V	M	Valine
B. N terminal at 118 location		
E118	WT	Glutamic Acid
E118D	M	Aspartic Acid
E118L	M	Leucine
E118N	M	Asparagine
C. C terminal at 130 location		
D130	WT	Aspartic Acid
D130E	M	Glutamic Acid
D130K	M	Lysine
D130V	M	Valine
D. C terminal at 133 location		
A133	WT	Alanine
A133D	M	Aspartic Acid
A133I	M	Isoleucine
A133M	M	Methionine
A133V	M	Valine

[0061] The observed gel shift behavior of Eg5 mutant proteins originates in altered detergent binding in the L5-modified proteins. A recent study (50) demonstrates that PAGE migration, SDS aggregation number, hydrodynamic radius and protein conformation are all directly linked. Observed variations in gel shifting in Figs. 1A-1D suggested that more than one type of modification in intra-protein contacts can be found in these mutant Eg5 motor domains and that affect the overall protein fold of the altered motor domains.

Example 3

Preservation of N-terminal local structure and C-terminal chemistry in the L5 loop are required for allostery and catalysis.

[0062] There are marked kinetic differences between the N-terminal and C-terminal substitutions of L5 loop residues. Compared with wildtype samples, single-site substitutions of E116 decreased the steady-state, basal ATP hydrolysis rate of Eg5 kinesin monomers (Fig. 2A, open boxes). However, mutants of E118 displayed increased activity (Fig. 2A, filled boxes). These results were consistent, whether or not the amino acid substitutions were synonymous with carboxylates. These data demonstrated that the native interactions of E116 and/or E118 in the motor domain influence ATP hydrolysis rates achievable by Eg5. Moreover, for mutants of these two positions in the L5 loop, the faithful difference in mutant kinetic behavior, despite conservative or pleiotropic changes, implicated the native sidechain of these N-terminal L5 residues not primarily in electrostatic interactions, but rather predominantly structural roles, likely propagating effects to the nucleotide site.

[0063] In contrast, the single-site C-terminal L5 substitutions did not have such uniform kinetic outcomes: rates of ATP hydrolysis for these residues were dependent on sidechain chemistry (Fig. 2B). For both D130 and A133, substitutions can result in either increased or decreased ATPase activity. D130E samples showed similar basal rates to wildtype Eg5 protein. Substitution of a positively-charged sidechain *in lieu* of aspartate at residue 130 decreased Eg5 activity, whereas replacement with a hydrophobic sidechain increased basal rates by approximately 50%. A133I and A133V Eg5 motor domains had ATP hydrolysis rates comparable to wildtype samples, but A133M was not as kinetically active. Only the substitution of a polar, charged sidechain for A133 produced an increase in enzymatic activity. Mutations, known to confer ispinesib-resistance (32,33), have opposing

changes in chemical character and have the greatest change in basal rates. Therefore, the chemical nature of these C-terminal L5 residues is key in their catalytic role.

[0064] The second type of perturbation to the L5 loop was use of allosteric small-molecule inhibitors. Specifically, the effects of two commonly-used inhibitors of Eg5, monastrol and S-trityl-L-cysteine were examined. Wildtype Eg5 had an IC_{50} of 8 and 4 μ M for monastrol and STC, respectively (black filled circles, Fig. 2C). Altering the chemistry of the E116, E118, D130, and A133, despite the paucity of direct contacts observed in crystal structures, changed the effect of small-molecule ligands on Eg5, encompassing the kinetic range from neutral to inhibited. For E118 substitutions (Fig. 2E), all variants of Eg5 were allosterically inhibited in equivalent measure to wildtype protein, suggesting that this carboxylate sidechain is not involved in allosteric inhibition.

[0065] In general, N-terminal mutations had a strong, positive correlation between loss of basal rate and loss of drug sensitivity, while C-terminal mutations did not have a clear relationship. All E116 (Fig. 2C), D130 (Fig. 2D), and A133 (Fig. 2F) mutants showed reduced sensitivity to monastrol and STC but to varying degrees. Only the E116R and D130K motor proteins displayed differing levels of inhibition between the two compounds used, the former being more sensitive to monastrol and the latter to STC. Of the C-terminal L5 mutants, D130V and A133D were the most insensitive to either inhibitor, as anticipated from studies on ispinesib (16). For the N-terminal L5 substitutions, loss of allosteric compound sensitivity was more marked in samples with aliphatic residue substitutions at positions 116 of the L5 loop, compared to Eg5 proteins with charged residues at these positions. Thus, the capacity for hydrogen-bond formation and the sidechain length of E116 are important for basal catalytic activity and for effective allosteric inhibition of Eg5.

Example 4

Secondary structural changes of L5-perturbed kinesins in solution can be quantitatively and rapidly measured by infrared spectroscopy.

[0066] To understand the role protein structure plays in allosteric interactions, conformational transitions throughout a globular protein need to be monitored in response to localized perturbations. There are no published structural descriptions of Eg5 mutants. Crystallographic reports of how allosteric small-compounds perturb wildtype Eg5•ADP complexes (19,20,34,35) note conformational changes in the L5 loop, but no pronounced changes in the nucleotide pocket or in other regions of the motor domain. Although X-ray

diffraction of protein crystals provides detailed atomic information, it can only provide a static single frame, or 'snapshot' of Eg5 inhibition.

[0067] Methods for determining the protein structure in solution can provide an ensemble conformational picture of polypeptides in a near-native environment. However, to date, there is no published method, rapid and quantitative, for screening conformational alterations resulting from peptide sequence variations that parallels high-throughput measurement of kinetic activity or genetic microarray profiling (36). Thus, there is a need to develop high-throughput methods to analyze protein structure in solution. Infrared spectroscopy is widely used as a technique for the analysis of protein secondary structure (37). The spectral region most commonly examined is the amide I band, centered at 1650 cm^{-1} , which arises from the delocalized C=O vibrations of the peptide linkage. Changes in the amide I region can be directly correlated with variation in the percentage of protein occupied in α -helix, β -sheet, or other structural motifs. Infrared data on Eg5 proteins in solution were obtained using reflectance spectral acquisition methods: spectra were collected using only 12-15 μL volumes and 20 min of time.

[0068] FTIR spectra were acquired on wildtype and all mutant Eg5 proteins alone and in the equimolar presence of monastrol or STC. Shown in Fig. 4A is a subset of the averaged spectral data on $^2\text{H}_2\text{O}$ -exchanged Eg5 proteins in the amide I' region ($1720\text{-}1600\text{ cm}^{-1}$). The normalized data share broadly similar lineshapes, arguing for similar overall secondary structure composition. Secondary structure contributions can be accurately measured through band narrowing and regression analysis (38,39) on averaged FTIR spectra (Fig. 3A-3C). Differences in amplitude for each frequency component measure proportionate changes in net secondary structure motifs between wildtype and mutant proteins. These quantitative changes are revealed after digital subtraction of the parent spectra, and can be decisively assigned to the gain or loss of structural motifs (Fig. 4B).

[0069] FTIR spectra on wildtype Eg5 proteins in the presence of allosteric inhibitors record structural changes that occur as a result of perturbing the L5 loop. Eg5•monastrol samples (Fig. 4B, top panel) have a 6% increased content for the 1633 cm^{-1} spectral component and a 8% decreased content for the 1628 cm^{-1} constituent, in comparison to uninhibited wildtype. From correlations of amide I' band frequency with secondary structural elements in proteins in $^2\text{H}_2\text{O}$ (40,41), both the 1628 cm^{-1} and 1633 cm^{-1} are assigned to low-frequency component of β -sheets. These changes are well above the level of noise: variation

from wildtype to wildtype sample was typically less than 1% at a given frequency (data not shown).

[0070] In contrast, the major changes between wildtype Eg5•STC and ligand-free wildtype samples are measured for the 1646 cm^{-1} and 1642 cm^{-1} spectral components, attributable to 9% increase in disordered structure and 10% decrease in 3_{10} helical content, respectively. Thus, interaction of wildtype Kinesin-5 protein with monastrol results in different spectral amide I' components than STC binding. Changes in the high-frequency beta sheet components at 1684, 1678, and 1672 cm^{-1} were minimal in both wildtype Eg5•STC and Eg5•monastrol bound samples. The 1665, 1657, and 1650 cm^{-1} modes, assignable to turns and alpha helices, show alterations in these structural elements between Eg5 bound with STC and bound with monastrol.

[0071] Alterations in peak height and shape of the broad amide I' envelope are detected also in the mutant Eg5 proteins (Fig. 4A). Difference spectra comparing mutant and wildtype data were obtained in a similar fashion to those described above, digitally subtracted and shown in Fig. 4B, middle and bottom panels. In general, E116 mutants have largest absorbance changes at 1646, 1642, 1633, and 1628 cm^{-1} and they show similar vibrational markers to STC-inhibited wildtype samples. (Fig. 4B, middle panel) E118 mutants have the largest changes in area at 1633 and 1628 cm^{-1} and they show similar vibrational markers to monastrol-inhibited wildtype samples. (Fig. 4B, bottom panel) As the overall pattern of secondary structural changes falls into either STC-like or monastrol-like spectrotypes, these data suggest two different populations of Eg5 conformers for these N-terminal L5 mutations. This interpretation is supported by the anomalous gel migration in Figs. 1A-1D. Thus, the N-terminal L5 substitutions result in modular changes in secondary structure of Eg5 in solution, irrespective of mutation or drug-binding.

[0072] Similar analyses were conducted for IR spectra collected from C-terminal L5 mutants in solution, in both their free- and ligand-bound forms. Substitution of A133 or D130 resulted in a greater number of possible conformational outcomes, as observed in the vibrational data (Figs. 5A and 5B). Overall, there was not a clear catalogue of conformational changes in the C-terminal L5 mutants, compared to the N-terminal L5 mutants. The C-terminal L5 mutants had the largest absorbance changes at 1646, 1642, 1633, and 1628 cm^{-1} and additional absorbance changes at 1665 and 1657 cm^{-1} (Fig. 5A). Some of these C-terminal A130 and D133 substitutions exhibited similar changes in secondary structural

content to the E116 mutants (Fig. 4B compared to Fig. 5B). However, other mutants in the C-terminus of the L5 loop had perturbations that are unique. Thus, as the C-terminal L5 mutants did not fall uniformly into the same spectrottype as STC- or monastrol-bound samples, these results suggested the long-range conformational effects are dependent on the chemical nature of the of the C-terminal residue sidechain.

Example 5

Exploratory principal component analysis identifies that N-terminal L5 mutants occupy a bimodal structural distribution in solution.

[0073] To create a model of how Eg5 undergoes allosteric conformational changes propagated from the L5 loop, a computationally-driven multivariate data analysis was used to extract structural information from the IR data. Techniques for deconvoluting multivariate data are common analytic tools in microarray data, systems biology studies, and chemometric detection (42-46), but its implementation for discovering atomic-level information is novel. The vibrational spectrum is fundamentally multivariate and complex in nature. Within a series of discrete measurements at defined frequency intervals found in each spectrum, the vast number of normal modes in the Kinesin-5 motor domain (>15,000 vibrational degrees of freedom) can have frequency overlap.

[0074] The initial foray into multivariate data analysis used principal component analysis (PCA). This non-parametric method that uses eigenvectors was applied to decompose vibrational spectra to a smaller set of computer-derived variables and to form models for predicting factors causing responses. Although PCA does not directly relate cause and effect, it serves to highlight the larger variations in the input data and find patterns in a set of spectra that on the surface can appear similar. Lastly, use of statistical algorithms to categorize the FTIR data provided an unbiased method that is mathematically grounded.

[0075] Spectroscopic data from the N-terminal and C-terminal L5 mutants were subjected to exploratory PCA. For proper formation of an initial model and for proper evaluation of the model prediction, cross-validation methods were not used, but rather the data was partitioned into two spectral data sets to be evaluated independently. By not using a sample before testing it, any bias was eliminated in the prediction accuracy (47). Wildtype Eg5 and the N-terminal L5 mutants served as the training set for PCA. The query, or test, set was the C-terminal L5 mutant spectra, used to verify and refine the predictive model. This

evaluation step, which included analysis of a second purified wildtype sample, was physically and temporally distinct from development of the initial training set.

[0076] The spectroscopic inputs are the band-narrowed amide I' frequencies from Eg5 protein in the absence of inhibitor, in the presence of monastrol, and in the presence of STC. In the outputs from these multivariate analyses (Figs. 6A-6C), the principal components (PC1 and PC2), or variables of the projection space, are linear combinations of the original frequencies.

[0077] Using wildtype and N-terminal L5 mutant Kinesin-5 proteins, distinct patterns were obtained from PCA. Of the ten-dimensional PCA output of IR frequencies projected on a two-dimensional plane formed by PC1 and PC2, four dominant vectors, defined by the 1642, 1646, 1633 and 1628 cm^{-1} IR frequencies, were found that accounted for 85% of the variance in the spectra in Fig. 6A. These patterns defined the amide I' frequencies associated with the E118 and E116 mutations and separated the motor proteins into two populations of conformers with high correlation. The first group contained E118 mutants and wildtype Eg5 (filled circles) in a tightly-clustered population. The second group was predominantly comprised of E116 mutants and the E116D-E118D double mutant (filled diamonds).

[0078] The 1642 and 1646 cm^{-1} vectors (dotted rays, Fig. 6A) were the most influential in distinguishing the two mutant populations. This principal component analysis emphasized the strong connection between residue substitution in the N-terminus of the L5 loop and changes in 3_{10} helical content and disordered structures, respectively. The 1633 and 1628 cm^{-1} vectors (Fig. 6A) also discriminated between the mutant populations, linking changes in Eg5 β -sheets with perturbations of the L5 loop.

[0079] The results from PCA were further supported by examination of additional protein samples that were not used in the initial model-building step. Independent PCA analysis of the C-terminal L5 mutants (Fig. 6B) also accounted for approximately 85% of the variance in these spectra. PCA of these test FTIR spectra (Fig. 6B) place the D130 and A133 mutants within identical structural populations of the N-terminal L5 mutants and predominantly fall into line with the wildtype/E118 data points. The fidelity of the training and test sets for this PCA experiment supports the conclusion that the initial model accurately predicted the conformer populations of Eg5 motor domains in which the L5 loop is challenged. Additionally, the high PCA correlation supports the conclusion that the curation

of Eg5 structures sampled a sufficient number of conformational possibilities to inform reasonable allosteric models.

[0080] In summary, spectroscopic signatures exhibited by drug-resistant variants were similar to those of wildtype human Eg5 (or KSP) inhibited by small-molecules. The above Eg5 variants are resistant to inhibition because their solution structure has already poised in the 'inhibited' state that may catalytically block the transition state. These convergent steady-state changes to the secondary structure of the Eg5 motor domain also supports the conclusion that the spectroscopic signature can serve as a biomarker for determining resistance to targeted chemotherapeutics in patient treatment. Thus, high-throughput measurement of modular conformation secondary structural changes within the human Eg5 motor domain permits conformational profiling of drug-resistant proteins.

Example 6

Monastrol and STC produce similar local, but dissimilar distal, structural changes.

[0081] Statistical algorithms were applied to infrared data to reveal features that were the basis for discriminating between the effects of monastrol and STC. The presence of monastrol or STC resulted in coincident, steady-state structural changes in mutant kinesin proteins (Fig. 6C): protein-inhibitor samples were separated by the 1646 and 1642 cm^{-1} components and fell into the same two PCA populations as free protein. This supports the conclusion that analogous modifications in disordered and 3_{10} helical content are reproducibly created in both Eg5 mutants and in allosteric inhibition.

[0082] The 1633 and 1628 cm^{-1} components differentiated the clustering of Eg5 samples incubated with monastrol or with STC (Fig. 6C). Regardless of which population the data points fell into, STC samples of the parent mutant generally had a greater 1628 cm^{-1} component (Fig. 6C, short dashed box), whereas monastrol samples had a greater 1633 cm^{-1} component (Fig. 6C, long dashed box). The 1633 and 1628 cm^{-1} frequencies arose from changes in β -sheet structures, which are found at the core of all kinesin motor domains. This PCA analysis underscored that, by IR spectroscopy, long-distance changes to β -sheets within the kinesin protein have been measured, changes that are a requirement for allostery. Moreover, the 5 cm^{-1} difference between the β -sheet components indicated that each frequency arose from a unique perturbation of this secondary structure. Thus, more than one means exist by which L5-initiated perturbations are coupled to distant secondary elements.

[0083] These observations by vibrational methods were corroborated by comparison of crystal structures of wildtype Eg5•ADP bound with monastrol (17) or STC. The 2.5 Å X-ray crystallographic determination of Eg5•ADP bound with a single STC molecule (Table 2) showed a canonical arrowhead-shaped globular structure. STC and monastrol were observed to stabilize similar overall states of the motor domain. The necklinker is positioned in the “docked” conformation seen in the monastrol-bound Eg5 structures, and the switch I and II regions also adopt conformations similar to those observed for bound monastrol.

Table 2. X-ray refinement statistics	
Resolution (Å)	2.5
$R_{\text{work}}/R_{\text{free}}$	0.241 / 0.275
Number of atoms	2749
Protein	2617
Ligand/ion	58
Water	74
B-factors	
Protein	25.537
Ligand/ion	26.120
Water	18.767
R.M.S deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.426

[0084] As expected from other crystallographic studies of Eg5 inhibitors, there were no large differences in the local structure of the allosteric site in the Eg5•ADP•STC structure. The overall fold of the L5 loop was similar to that of two published monastrol-bound Eg5 structures. However, STC binding resulted in a slightly greater restriction of the allosteric site by the L5 loop: STC binding occludes approximately 80 Å² more of the Eg5 surface from solvent access than does monastrol. Of the residues examined in this study, only the rotamer configuration of E118 in Eg5•ADP•STC differed slightly from monastrol-bound Eg5.

[0085] Long-range structural effects were shown in the β-sheet core of the motor domain, due to binding of L5-directed inhibitors. The central β-sheet structure of all kinesin motor domains consists of a parallel β-sheet, with a topology (or β-strand order) of 2183, adjacent to an antiparallel β-sheet with a topology of 7645. There are clear distortions of the antiparallel β-sheet that likely impose novel strain on the motor domain core in presence of these allosteric inhibitors.

[0086] Strands that undergo the largest changes are β4, β5, β6, and β7. Although all strands typically exhibit a right-handed twist in higher-order structures, the N-terminus of the β4 strand is pulled closer to the α2 helix in both STC- and monastrol-bound structures, compared to the native Eg5 structure. The β5 strand is not formed in the Eg5•STC structure. The β6 and β7 strands have large distortions that differ in magnitude and direction, dependent on whether STC or monastrol was bound to the L5 loop. Direct comparison of the STC- and monastrol-bound Eg5 structures stresses the divergence in the twist of the β6 and β7 strands. Exaggeration in the twist of the β6 and β7 strands are also observed in other structures of Eg5 co-crystallized with L5-directed inhibitors.

Example 7

A high prediction accuracy exists between IR and X-ray diffraction determination of Eg5 helical structures.

[0087] Current IR data did not inherently contain information on the precise site(s) of origin within the Eg5 motor domain. Inspection of relevant crystal structures can facilitate the further assignment of an IR signature to a specific protein segment. Examination of fractional secondary structure content of wildtype Eg5 motor proteins determined from X-ray diffraction methods supported the multivariate findings of Kinesin-5 in solution. From DSSP analysis (48) of Eg5 structures in the PDB, increased 3₁₀ helix content was observed when comparing ligand-bound versus free Eg5 proteins (Table 3). In Eg5•ADP crystal structures

(22), four 3_{10} helices were found within the $\alpha 3$, $\alpha 5$, and $\alpha 6$ helices and in the switch I loop. Thus, 3.3% of residues in the Eg5 motor domain were found in 3_{10} helices in Eg5•ADP crystal structures. In inhibitor-complexed Eg5 proteins, there was generally a gain of two 3_{10} helices in the L5 loop, increasing the fractional content of 3_{10} helices to 4.9%

Table 3. Comparison of fractional secondary structure composition in inhibitor-free and inhibitor-bound crystal structures of Eg5. Secondary structure assignments in crystallographic data are obtained from DSSP analysis. Fractional compositions were calculated for each protein within the asymmetric unit cell and averaged, when appropriate. Subtraction of inhibitor-bound values from inhibitor-free Eg5 structures is shown.

Compared structures	% change in secondary structure					
	β -sheet	α -helix	3_{10} helix	turn	bend	β -bridge
3KEN – 1II6 (STC)	-0.7	1.5	0.0	-1.5	0.8	0.0
1Q0B – 1II6 (mon)	3.0	-1.4	1.6	0.3	-0.7	0.0
1X88 – 1II6 (mon)	2.3	-1.5	1.6	0.9	-1.6	0.0
1YRS – 1II6	0.0	-2.4	1.2	1.2	0.9	-0.3
2IEH – 1II6	2.7	-2.6	0.0	2.4	0.7	-0.3

[0088] Although the IR spectrum is a reflection of the motor protein ensemble, it is believed that the above IR-detectable changes at 1642 and 1646 cm^{-1} are localized to the L5 loop, as it is the only protein segment that shows a conformational isomerization between 3_{10} helices and a random loop. Moreover, this stringent comparison between the solution structural determinations and crystallographic data calls attention to the sensitivity of the vibrational methods employed here: even changes in 6 residues can be accurately detected within the amide I' envelope.

Example 8

These IR spectrotypes serve as novel structural biomarkers to predict drug sensitivity.

[0089] The above experiments demonstrate that inhibitor efficacy can be measured structurally: it is predicted that a large 1628 cm^{-1} component would be observed in PCA of

more potent inhibitors of Eg5. The uncovered interactions that are essential in determining drug sensitivity in Eg5 mutants *in vitro* can be extended to predicting how human polymorphisms can affect clinical responsiveness.

[0090] Herein particular structural signatures were demonstrated to be correlated with inhibitor resistance in Eg5 variants. The first condition for inhibitor resistance in Eg5 is the conformational selection of the L5 loop in solution. For the N-terminal L5 substitutions, there was a clear difference in the pattern of secondary structure changes observed in the E116 mutant proteins when compared to those of the E118 substitutions (Figs. 4A, 4B, and 6A). The kinetic sensitivity of E118 mutants to drug inhibition (Fig. 2A and 2E) is correlated with adoption of 3_{10} helical structure in the L5 loop and a large 1642 cm^{-1} vector (Fig. 6A). In line with this conclusion, the clustering of the E116 mutant samples in PCA indicated that a large 1646 cm^{-1} component was strongly correlated with drug resistance. Thus, loss of the 3_{10} helix in the L5 loop unilaterally prohibits binding of small molecules to the allosteric site, as shown by the D130V mutation that confers resistance in Eg5 to the clinical trial drugs in purified form, in tumor cell culture, and in xenograft model systems (32,33).

[0091] Another component in inhibitor resistance in Kinesin-5 proteins is the mechanochemical state of the antiparallel β -sheet. For Eg5 variants that populate the 1646 cm^{-1} cluster in the principal component analysis, the lack of 3_{10} helix functionally overrides any conformational changes observed in the β -sheet. Therefore, there is a hierarchy of structural changes that dictate inhibitor resistance. For Eg5 variants that populate the 1642 cm^{-1} cluster in PCA, only genotypes that result in strong 1633 cm^{-1} vectors remain sensitive to drug inhibition. If there is a twist of the β_6 and β_7 strands that resembles the STC-bound Eg5 and thus exhibits a strong 1628 cm^{-1} component in multivariate analysis (Fig. 6B), this torque on the antiparallel β -sheet negates drug sensitivity, even if 3_{10} helices are adopted. This is the case for the established ispinesib-resistant mutant, A133D.

[0092] From this work, it is believed that isomerization of the L5 loop and distortions of the central β -sheet can also serve as novel structural biomarkers to predict drug sensitivity of Eg5. This biomarker provides correlation between protein structure and therapeutic outcome, which is advantageous because the final target for clinical drugs is typically polypeptides. Case in point, more than half of the mutants examined had catalytic rates comparable or greater than wildtype samples. Seemingly normal Eg5 kinesin function could harbor a polymorphism that would confer resistance to clinical drugs, and this biomarker

could discriminate against Eg5-directed inhibitors. This is a rapid and quantitative method for assaying conformational alterations, resulting from peptide sequence variations, of protein ensembles in solution. This biophysical approach answers 'how much' an individual protein variant is resistant to a drug. It is believed that these structural biomarkers based on conformational changes due to the sequence polymorphism can be used in the clinic for appropriate treatment options for kinesin-related diseases.

[0093] Quantitative measurement of structural changes as a function of inhibitor concentration will provide preliminary data toward establishing correlations between the relative amount of structural changes invoked by a polymorphism and the concentration of a compound necessary to inhibit the variant protein.

[0094] It is believed that the use of structural biomarkers is extendable to other protein targets for cancer drugs. The tyrosine kinase, BCR-ABL, will be tested with these methods, as it has the greatest molecular, biochemical, and structural information available and there are two generations of inhibitors in current clinical use. The BCR-ABL oncogene, which is the product of the Philadelphia chromosome 22q, encodes a chimeric BCR-ABL protein that has constitutively activated ABL tyrosine kinase activity and is the underlying cause of chronic myeloid leukemia (CML). The impact of Gleevec, a small molecule inhibitor of the oncogenic form of this protein, on the field of cancer therapy has been dramatic. This drug is the current gold standard of targeted chemotherapy, as it has not only changed how newly diagnosed patients with CML are treated, but it also improves their prognosis. Gleevec resistance can develop in patients, typically caused by point mutations to the kinase domain that reduce binding of Gleevec by either direct or indirect mechanisms. This structural biomarker can be applied to purified proteins from patient samples. This method can be used to locate a structural biomarker for BCR-ABL and offer new information on mutants that ablate Gleevec binding by indirect mechanisms and for which no reason for resistance has been established.

[0095] Lastly, the finding and describing of fully-developed structural biomarkers can predict drug efficacy in patients. Future treatment strategies will be based on an analysis of an individual's cancer cells or other cells, and proteins from those cancer cells or other cells, which would then allow development of a profile of likely sensitivity of the proteins to targeted therapies. Based on this information, each individual might then be assigned to a combination regimen that best matches the profile from the cancer cells or other cells.

Toward this treatment strategy, initial efforts will examine documented medication response profiles and somatic polymorphisms of patients in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. The SEER registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and outcome data (survival). In addition, the Residual Tissue Repository (RTR) program provides representative sampling of biospecimens associated with SEER patients for testing. Somatic polymorphisms can be determined for patients who were treated with tyrosine kinase inhibitors. When kinesin spindle protein (KSP) inhibitors are adopted into standard clinical practice, parallel experiments can be performed as well.

[0096] Clinical trials will evaluate whether oncologists use of structural biomarkers improves therapeutic response in patients. In the process of achieving a more individualized treatment strategy, this clinical trial will evaluate the performance of 'random' selection of agents (current practice), genomic signature-based selection, structural biomarker selection, or a combination of genomic- and spectroscopic signatures for selection. This would evaluate the predictive value of the different types of biomarkers toward therapeutic response in the hands on physicians.

REFERENCES

1. Gunasekaran, K., Ma, B., and Nussinov, R. (2004) *Proteins* **57**, 433-443
2. Monod, J., Wyman, J., and Changeux, J. P. (1965) *Journal of molecular biology* **12**, 88-118
3. Kern, D., and Zuiderweg, E. R. (2003) *Current opinion in structural biology* **13**, 748-757
4. Cooper, A., and Dryden, D. T. (1984) *Eur Biophys J* **11**, 103-109
5. Hotha, S., Yarrow, J. C., Yang, J. G., Garrett, S., Renduchintala, K. V., Mayer, T. U., and Kapoor, T. M. (2003) *Angewandte Chemie (International ed)* **42**, 2379-2382
6. DeBonis, S., Skoufias, D. A., Lebeau, L., Lopez, R., Robin, G., Margolis, R. L., Wade, R. H., and Kozielski, F. (2004) *Mol Cancer Ther* **3**, 1079-1090
7. Sunder-Plassmann, N., Sarli, V., Gartner, M., Utz, M., Seiler, J., Huemmer, S., Mayer, T. U., Surrey, T., and Giannis, A. (2005) *Bioorg Med Chem* **13**, 6094-6111
8. Maliga, Z., Kapoor, T. M., and Mitchison, T. J. (2002) *Chemistry & biology* **9**, 989-996
9. DeBonis, S., Simorre, J. P., Crevel, I., Lebeau, L., Skoufias, D. A., Blangy, A., Ebel, C., Gans, P., Cross, R., Hackney, D. D., Wade, R. H., and Kozielski, F. (2003) *Biochemistry* **42**, 338-349
10. Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L., and Mitchison, T. J. (1999) *Science* **286**, 971-974
11. Brier, S., Lemaire, D., Debonis, S., Forest, E., and Kozielski, F. (2004) *Biochemistry* **43**, 13072-13082
12. Cox, C. D., Breslin, M. J., Mariano, B. J., Coleman, P. J., Buser, C. A., Walsh, E. S., Hamilton, K., Huber, H. E., Kohl, N. E., Torrent, M., Yan, Y., Kuo, L. C., and Hartman, G. D. (2005) *Bioorg Med Chem Lett* **15**, 2041-2045
13. Cochran, J. C., Gatial, J. E., 3rd, Kapoor, T. M., and Gilbert, S. P. (2005) *The Journal of biological chemistry* **280**, 12658-12667
14. Skoufias, D. A., DeBonis, S., Saoudi, Y., Lebeau, L., Crevel, I., Cross, R., Wade, R. H., Hackney, D., and Kozielski, F. (2006) *The Journal of biological chemistry* **281**, 17559-17569
15. Lad, L., Luo, L., Carson, J. D., Wood, K. W., Hartman, J. J., Copeland, R. A., and Sakowicz, R. (2008) *Biochemistry* **47**, 3576-3585
16. Luo, L., Parrish, C. A., Nevins, N., McNulty, D. E., Chaudhari, A. M., Carson, J. D., Sudakin, V., Shaw, A. N., Lehr, R., Zhao, H., Sweitzer, S., Lad, L., Wood, K. W., Sakowicz, R., Annan, R. S., Huang, P. S., Jackson, J. R., Dhanak, D., Copeland, R. A., and Auger, K. R. (2007) *Nat Chem Biol* **3**, 722-726
17. Maliga, Z., and Mitchison, T. J. (2006) *BMC Chem Biol* **6**, 2
18. Yan, Y., Sardana, V., Xu, B., Homnick, C., Halczenko, W., Buser, C. A., Schaber, M., Hartman, G. D., Huber, H. E., and Kuo, L. C. (2004) *Journal of molecular biology* **335**, 547-554
19. Fraley, M. E., Garbaccio, R. M., Arrington, K. L., Hoffman, W. F., Tasber, E. S., Coleman, P. J., Buser, C. A., Walsh, E. S., Hamilton, K., Fernandes, C., Schaber, M. D., Lobell, R. B., Tao, W., South, V. J., Yan, Y., Kuo, L. C., Prueksaritanont, T., Shu, C., Torrent, M., Heimbrook, D. C., Kohl, N. E., Huber, H. E., and Hartman, G. D. (2006) *Bioorg Med Chem Lett* **16**, 1775-1779
20. Garbaccio, R. M., Fraley, M. E., Tasber, E. S., Olson, C. M., Hoffman, W. F., Arrington, K. L., Torrent, M., Buser, C. A., Walsh, E. S., Hamilton, K., Schaber, M. D., Fernandes, C., Lobell, R. B., Tao, W., South, V. J., Yan, Y., Kuo, L. C., Prueksaritanont, T., Slaughter, D. E., Shu, C., Heimbrook, D. C., Kohl, N. E., Huber, H. E., and Hartman, G. D. (2006) *Bioorg Med Chem Lett* **16**, 1780-1783

21. Kaan, H. Y., Ulaganathan, V., Hackney, D. D., and Kozielski, F. (2009) *Biochem J*
22. Turner, J., Anderson, R., Guo, J., Beraud, C., Fletterick, R., and Sakowicz, R. (2001) *The Journal of biological chemistry* **276**, 25496-25502
23. Cochran, J. C., and Gilbert, S. P. (2005) *Biochemistry* **44**, 16633-16648
24. Bodey, A. J., Kikkawa, M., and Moores, C. A. (2009) *Journal of molecular biology* **388**, 218-224
25. Parke, C. L., Wojcik, E. J., Kim, S., and Worthylake, D. K. (2009) *The Journal of biological chemistry* **285**, in press
26. Maliga, Z., Xing, J., Cheung, H., Juszczak, L. J., Friedman, J. M., and Rosenfeld, S. S. (2006) *The Journal of biological chemistry* **281**, 7977-7982
27. Brier, S., Lemaire, D., DeBonis, S., Kozielski, F., and Forest, E. (2006) *Rapid Commun Mass Spectrom* **20**, 456-462
28. Wojcik, E. J., Dalrymple, N. A., Alford, S. R., Walker, R. A., and Kim, S. (2004) *Biochemistry* **43**, 9939-9949
29. Deavours, B. E., Reddy, A. S., and Walker, R. A. (1998) *Cell Motil Cytoskeleton* **40**, 408-416
30. Learman, S. S., Kim, C. D., Stevens, N. S., Kim, S., Wojcik, E. J., and Walker, R. A. (2009) *Biochemistry* **48**, 1754-1762
31. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr A* **47 (Pt 2)**, 110-119
32. Blagden, S. P., Molife, L. R., Seebaran, A., Payne, M., Reid, A. H., Protheroe, A. S., Vasist, L. S., Williams, D. D., Bowen, C., Kathman, S. J., Hodge, J. P., Dar, M. M., de Bono, J. S., and Middleton, M. R. (2008) *Br J Cancer* **98**, 894-899
33. Carol, H., Lock, R., Houghton, P. J., Morton, C. L., Kolb, E. A., Gorlick, R., Reynolds, C. P., Maris, J. M., Keir, S. T., Billups, C. A., and Smith, M. A. (2009) *Pediatr Blood Cancer* **53**, 1255-1263
34. Garcia-Saez, I., DeBonis, S., Lopez, R., Trucco, F., Rousseau, B., Thuery, P., and Kozielski, F. (2007) *The Journal of biological chemistry* **282**, 9740-9747
35. Zhang, B., Liu, J. F., Xu, Y., and Ng, S. C. (2008) *Biochem Biophys Res Commun* **372**, 565-570
36. Banci, L., Bertini, I., Cusack, S., de Jong, R. N., Heinemann, U., Jones, E. Y., Kozielski, F., Maskos, K., Messerschmidt, A., Owens, R., Perrakis, A., Poterszman, A., Schneider, G., Siebold, C., Silman, I., Sixma, T., Stewart-Jones, G., Sussman, J. L., Thierry, J. C., and Moras, D. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 1208-1217
37. Barth, A., and Zscherp, C. (2002) *Q Rev Biophys* **35**, 369-430
38. Goormaghtigh, E., De Meutter, J., Vanloo, B., Brasseur, R., Rosseneu, M., and Ruyschaert, J. M. (1989) *Biochimica et biophysica acta* **1006**, 147-150
39. Hutchison, R. S., Betts, S. D., Yocum, C. F., and Barry, B. A. (1998) *Biochemistry* **37**, 5643-5653
40. Susi, H., and Byler, D. M. (1986) *Methods in Enzymology* **130**, 290-311
41. Surewicz, W. K., Mantsch, H. A., and Chapman, D. (1993) *Biochemistry* **32**, 389-394
42. Misra, J., Schmitt, W., Hwang, D., Hsiao, L. L., Gullans, S., and Stephanopoulos, G. (2002) *Genome Res* **12**, 1112-1120
43. Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., and Drineas, P. (2007) *PLoS Genet* **3**, 1672-1686
44. Bylesjo, M., Eriksson, D., Kusano, M., Moritz, T., and Trygg, J. (2007) *Plant J* **52**, 1181-1191
45. Lin, Z., and Altman, R. B. (2004) *Am J Hum Genet* **75**, 850-861

46. Beckwith-Hall, B. M., Nicholson, J. K., Nicholls, A. W., Foxall, P. J., Lindon, J. C., Connor, S. C., Abdi, M., Connelly, J., and Holmes, E. (1998) *Chem Res Toxicol* **11**, 260-272
47. Dupuy, A., and Simon, R. M. (2007) *J Natl Cancer Inst* **99**, 147-157
48. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577-2637
49. Krimm, S., and Bandekar, J. (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. in *Advances in Protein Chemistry* (Anfinsen, C. B., Edsall, J. T., and Richards, F. M. eds.), Academic Press, New York. pp 181-364
50. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009) *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1760-1765
51. Jin, M., Song, G., Carman, C. V., Kim, Y. S., Astrof, N. S., Shimaoka, M., Wittrup, D. K., and Springer, T. A. (2006) *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5758-5763
52. Hirose, K., Akimaru, E., Akiba, T., Endow, S. A., and Amos, L. A. (2006) *Mol Cell* **23**, 913-923
53. Coureux, P. D., Wells, A. L., Menetrey, J., Yengo, C. M., Morris, C. A., Sweeney, H. L., and Houdusse, A. (2003) *Nature* **425**, 419-423
54. Reubold, T. F., Eschenburg, S., Becker, A., Kull, F. J., and Manstein, D. J. (2003) *Nature structural biology* **10**, 826-830
55. Pasinelli, P & Brown, R. (2006) *Nature Reviews Neuroscience* **7**, 710-723.
56. Schiffman, D., Shasman, D.I., Zee, R., Iakoubova, O., Louie, J., Devlin, J., and Ridker, P. (2008) *Journal of the American College of Cardiology* **51**, 444-448
57. Iakoubova, O., Sabatine, M., Rowland, C., Rong, C., Catanese, J., Ranade, K., Simonsen, Kirchgessner, R., Cannon, C., Devlin, J., Braunwald, E. (2008) *Journal of the American College of Cardiology* **51**, 449-455
58. Iakoubova, O., Tong, C., Rowland, C., Kirchgessner, T., Young, B., Arellano, A., Shiffman, D., Sabatine, M., Campos, H., Packard, C., Pfeffer, M., White, T., Braunwald, E., Shepherd, J., Devlin, J., Sacks, F. (2008) *Journal of the American College of Cardiology* **51**, 435-443
59. Chene, P. (2002) *Nature Reviews, Drug Discovery* **1**, 665-673
60. Hochhaus, A., Kreil, S., Corbin, A.S., La Rosee, P., Muller, M.C., Lahaye, T., Hanfstein, B., Schoch, C., Cross, N.C.P., Berger, U., Gschaidmeier, H., Druker, B.J., and Hehlmann, R. (2002) *Leukemia* **16**, 2190-96
61. Shah, N.P., Nicoll, J.M., Nagar, B., Gorre, M.E., Paquette, R.L., Kuriyan, J., and Sawyers, C.L. (2002) *Cancer Cell* **2**, 117-125
62. Carter, B.Z., Mak, D.H., Shi, Y., Schober, W.D., Wang, R.-Y., Konopleva, M., Koller, E., Dean, N.M., and Andreeff, M. (2006) *Cell Cycle* **5**, 2223-2229

[0097] The complete disclosures of all references cited in this specification are hereby incorporated by reference. Also incorporated by reference are the complete disclosures of the following: Provisional application no. 61/154,121, filed February 20, 2009, entitled "Protein Structural Biomarkers to Guide Targeted Chemotherapies"; Provisional application no. 61/261,939, filed November 17, 2009, entitled "Protein Structural Biomarkers to Guide Targeted Chemotherapies"; R. Buckley *et al.*, "Conformational specificity in allosteric signaling: high-throughput measurement of modular secondary structural changes within human Eg5 kinesin," an abstract submitted to the Biophysical Society 53rd Annual Meeting published online February 2009; B. Jun and S. Kim, "Millisecond time-lapsed monitoring of ATP hydrolysis by human Eg5 kinesin: real-time dynamics of conformation and chemistry *in vitro*," an abstract submitted to the Biophysical Society 53rd Annual Meeting published online February 2009; C.L. Parke *et al.*, "Crystal structure of HsEg5 in complex with S-trityl-L-cysteine," an abstract submitted to the Biophysical Society 53rd Annual Meeting published online by February 2009; E. D. Kim *et al.*, "Allosteric drug discrimination is coupled to mechanochemical changes in the Kinesin-5 motor core," a manuscript accepted by the Journal of Biological Chemistry in February 2010; and B. Jun and S. Kim, "Real-Time structural transitions are coupled to chemical steps in ATP hydrolysis by Eg5 kinesin," a manuscript accepted by the Journal of Biological Chemistry in February 2010. In the event of an otherwise irreconcilable conflict, however, the present specification shall control.

What is claimed:

1. A method for predicting the probable response of a patient to a drug compound; wherein the patient has a disease; wherein the drug compound inhibits a protein that is associated with the disease; and wherein the protein is polymorphic in the population, such that there is substantial variation in the responsiveness of the disease to the drug compound between individuals as a result of such polymorphism; said method comprising the steps of:

(a) determining whether the variant of the protein expressed by the patient has greater spectroscopic similarity to: (i) a variant of the protein that is known to be responsive to the drug compound, in the absence of the drug compound, or (ii) a complex of the known-responsive-variant of the protein, when bound to the drug compound; and

(b) predicting that the patient will likely be responsive to the drug compound if there is greater spectroscopic similarity to (i); or predicting that the patient will likely be unresponsive or less responsive to the drug compound if there is greater spectroscopic similarity to (ii);

wherein the degree of vibrational spectroscopic similarity is determined as absorption or transmittance at one or more wavelengths at which there is a significant difference in absorption or transmittance between (i) and (ii).
2. The method of Claim 1, wherein step (a) is conducted by direct vibrational spectroscopic measurement.
3. The method of Claim 2, wherein step (a) is conducted by infrared or Raman spectroscopic measurement.
4. The method of Claim 2, wherein step (a) is conducted by Fourier-transform infrared spectroscopy.
5. The method of Claim 1, wherein step (a) is conducted by genetic analysis of the patient's DNA or mRNA encoding the protein, and comparing the patient's allele or alleles for the protein to those of one or more reference samples for which direct spectroscopic measurements of the protein have previously been conducted.

6. The method of Claim 1, wherein the disease is a cancer.
7. The method of Claim 1, wherein the drug compound is a chemotherapeutic agent.
8. The method of Claim 7, wherein the drug compound is a chemotherapeutic agent that acts as an allosteric inhibitor of the protein.
9. The method of Claim 8, wherein the drug compound inhibits mitosis.
10. The method of Claim 1, wherein the protein is an ATPase.
11. The method of Claim 1, wherein the ATPase is selected from the group consisting of muscle myosin II, Na^+, K^+ -ATPase, P glycoprotein, H^+, K^+ - pump, vacuolar H^+ -ATPase, and DNA topoisomerase II.
12. The method of Claim 1, wherein the protein is a kinesin selected from the group consisting of Eg5, Kif5A, Kif21A, Kif1B, and Kif6.
13. The method of Claim 1, wherein the protein is kinesin Eg5, and wherein the drug compound inhibits kinesin Eg5.
14. The method of Claim 13, wherein the drug is selected from the group consisting of monastrol, S-trityl-L-cysteine and its derivatives, quinazolines (e.g., ispinesib), adociasulfates, tetrahydroishquinolines, dihydropyrazoles, thiophenes, pyrrolotriazines, thiazoles, gossypol, indoles, and biphenyls.
15. The method of Claim 13, wherein the drug is monastrol.
16. The method of Claim 13, wherein the drug is S-trityl-L-cysteine.
17. The method of Claim 1, wherein the protein is the tyrosine kinase BCR-ABL.
18. The method of Claim 17, wherein the drug is Gleevec.

Fig. 1A

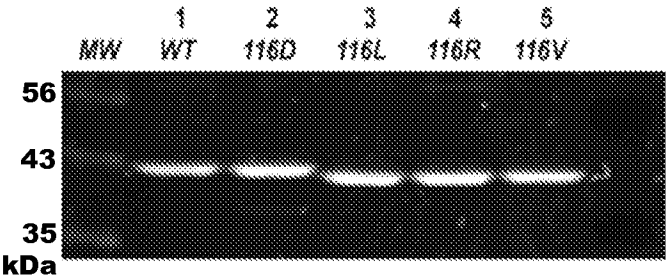


Fig. 1B

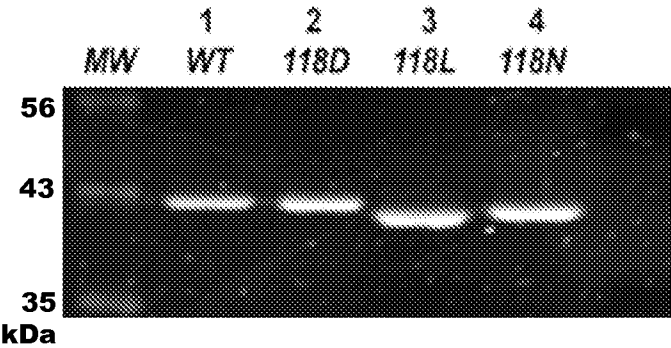


Fig. 1C

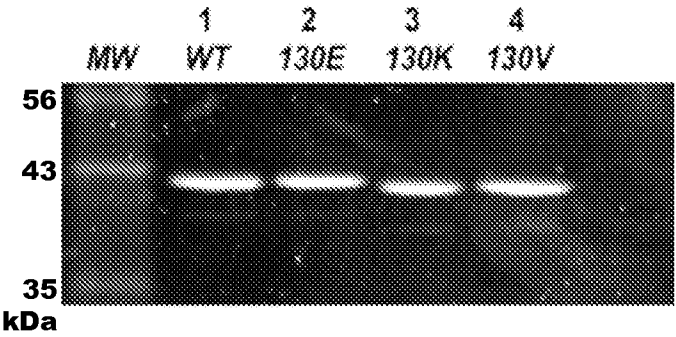
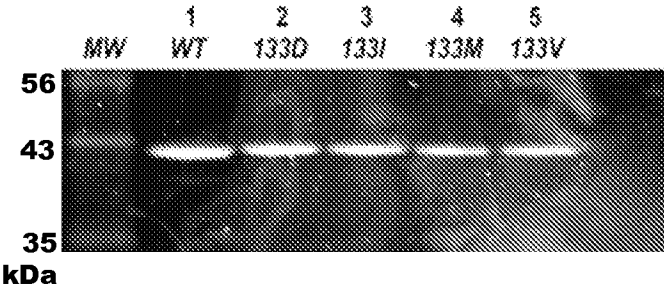


Fig. 1D



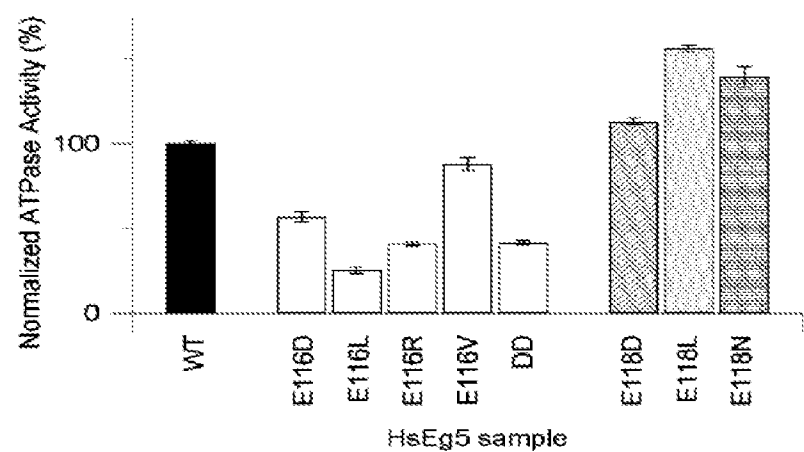


Fig. 2A

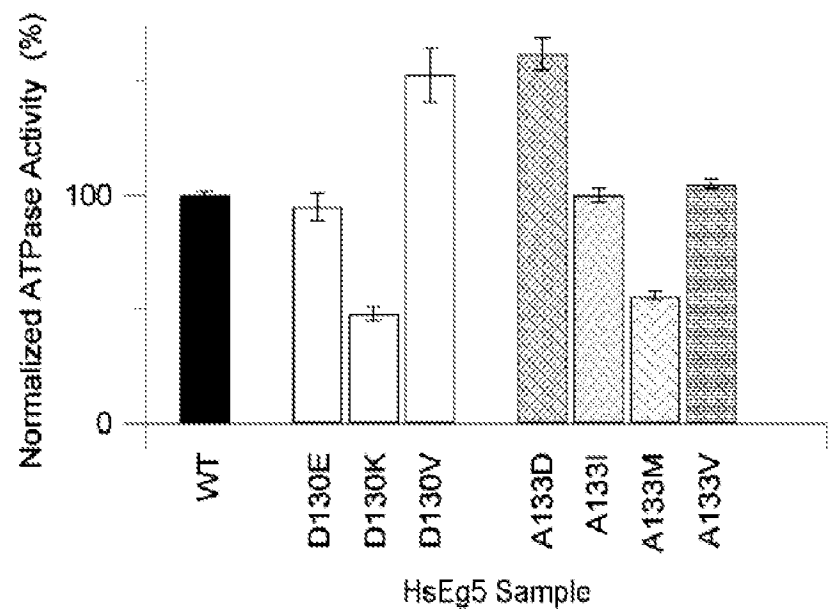


Fig. 2B

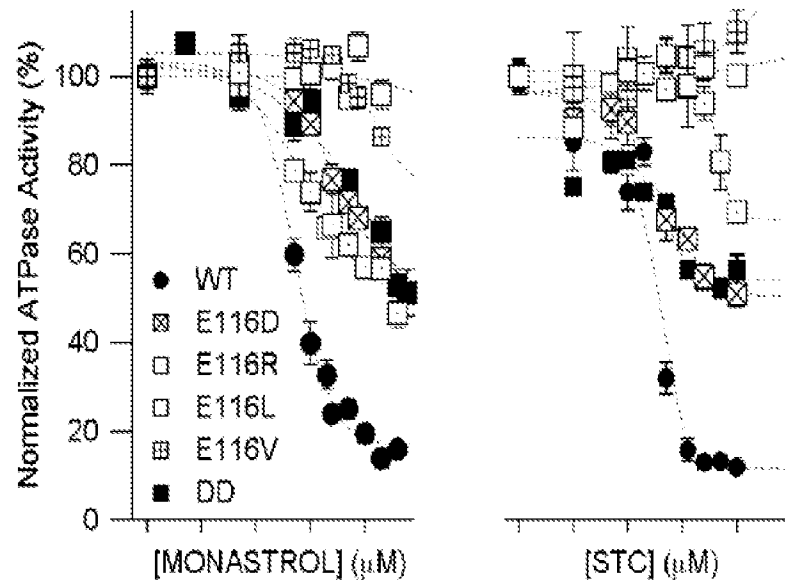


Fig. 2C

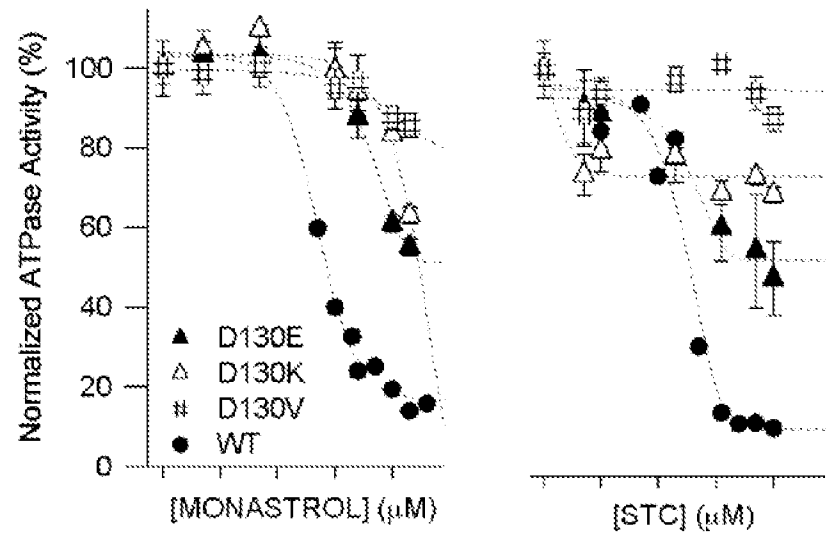


Fig. 2D

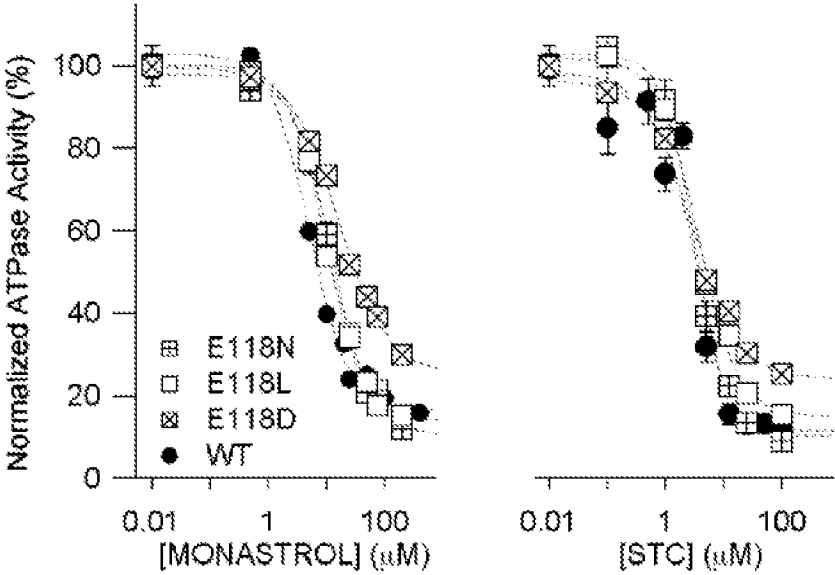


Fig. 2E

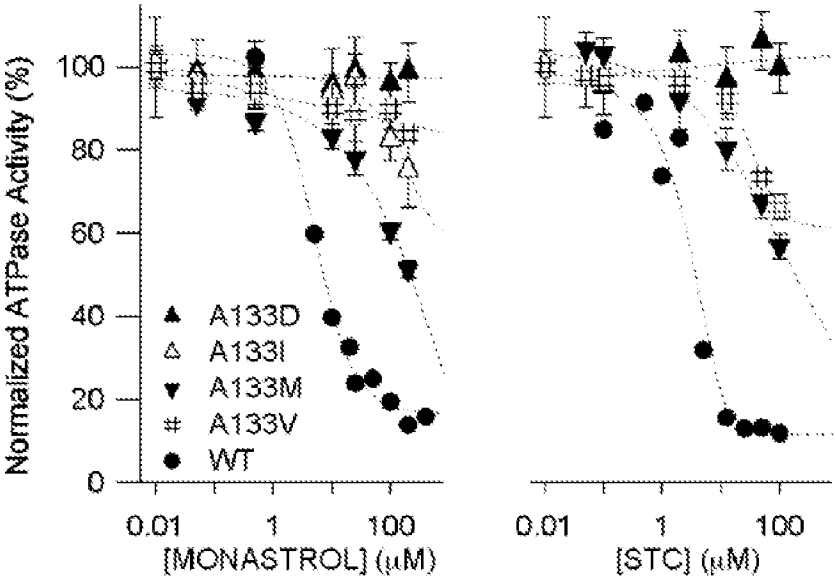


Fig. 2F

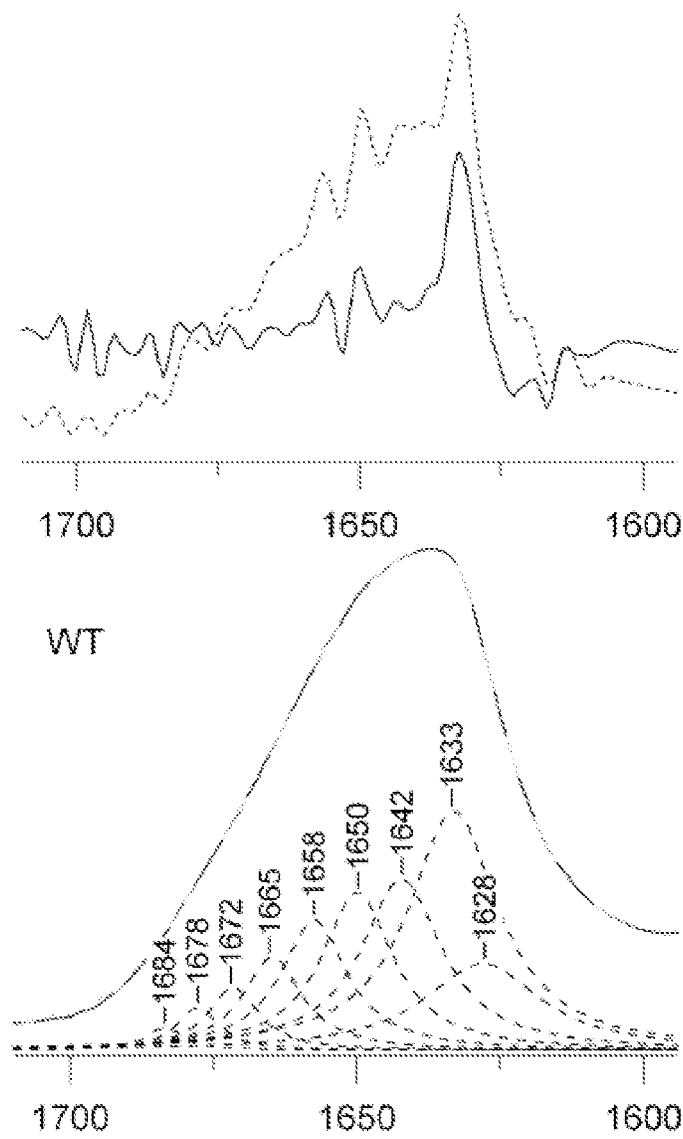


Fig. 3A

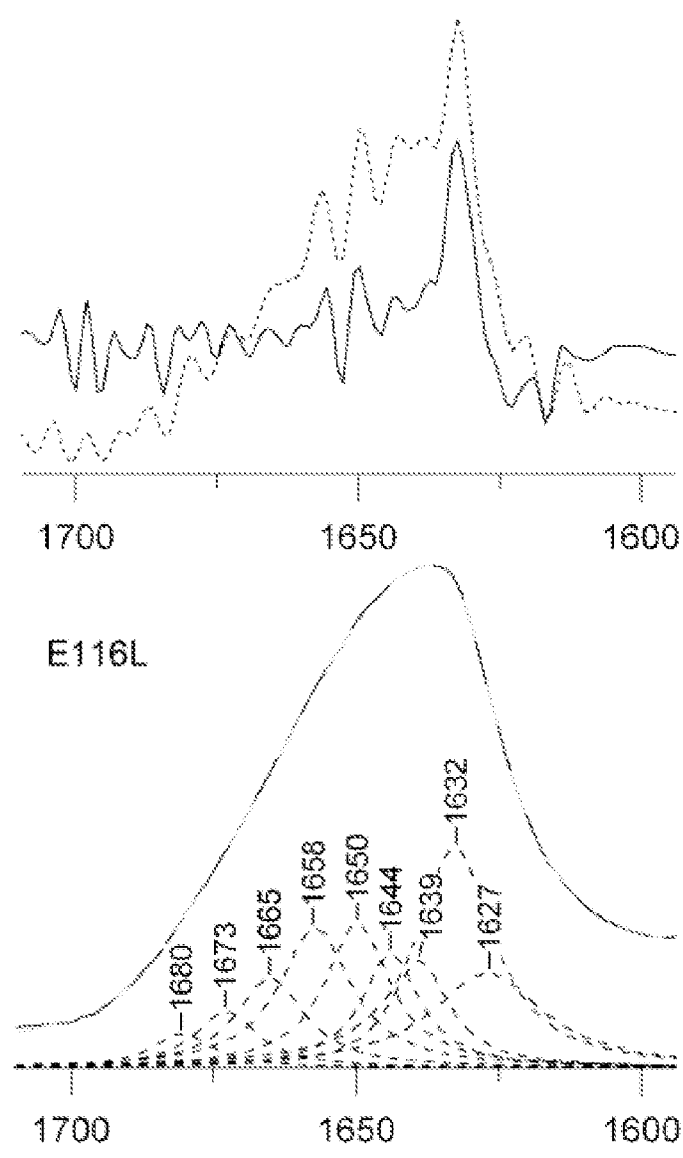


Fig. 3B

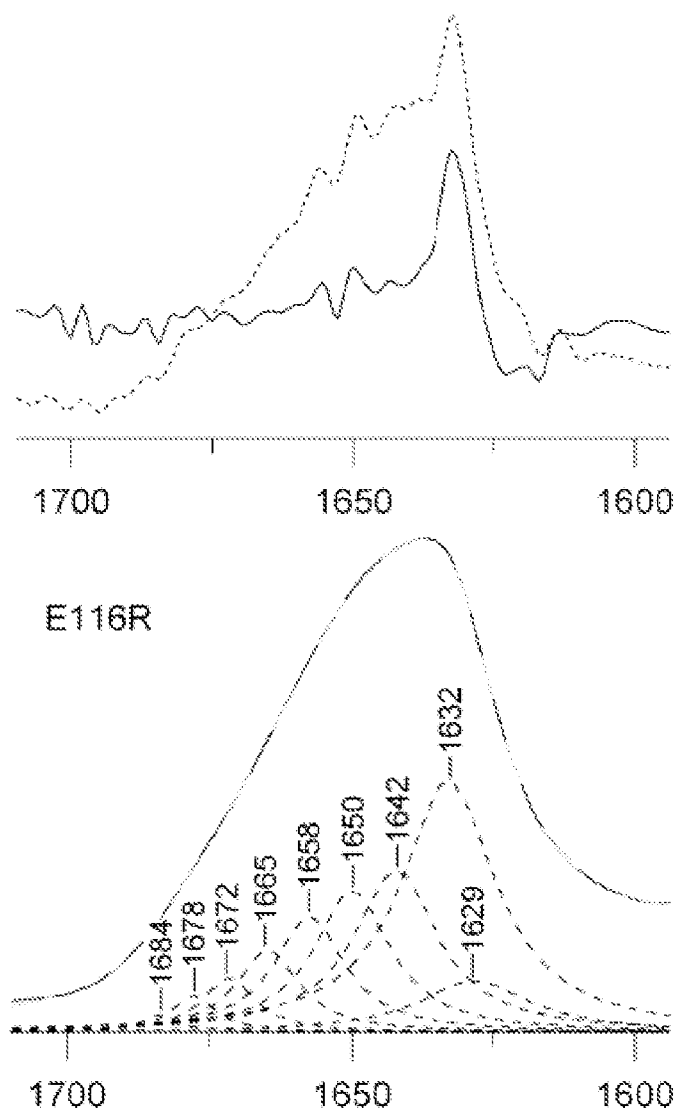
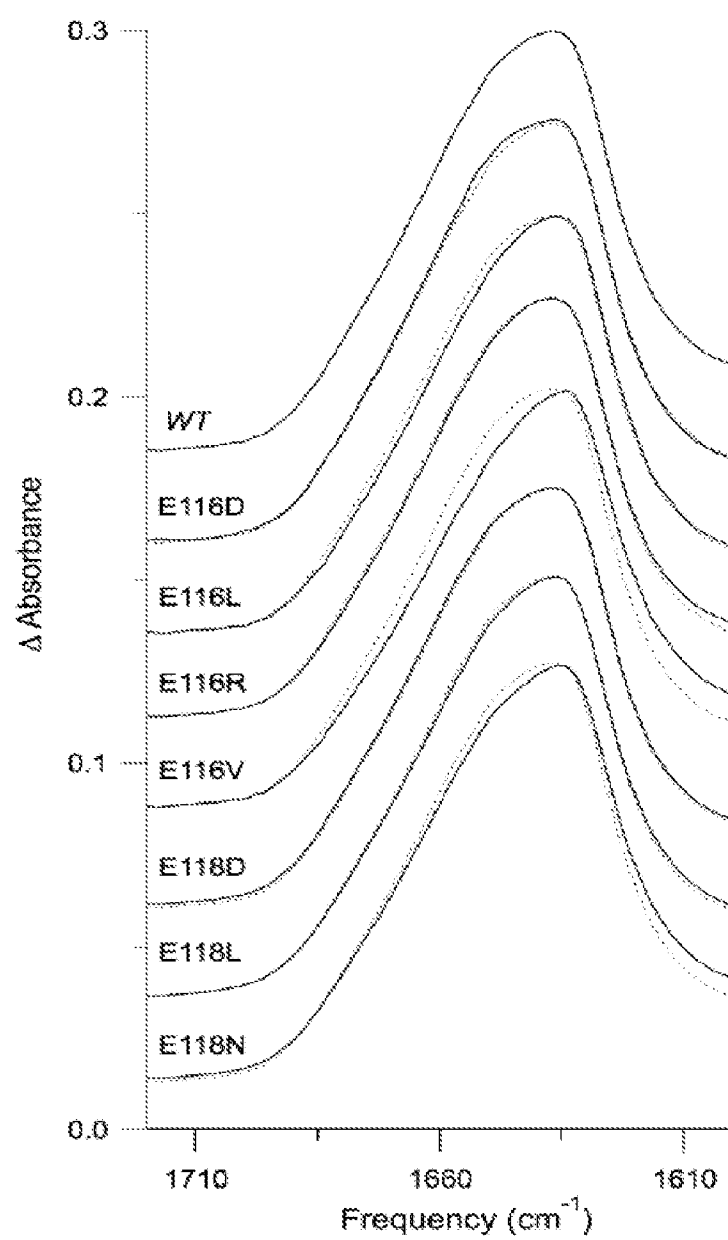


Fig. 3C

**Fig. 4A**

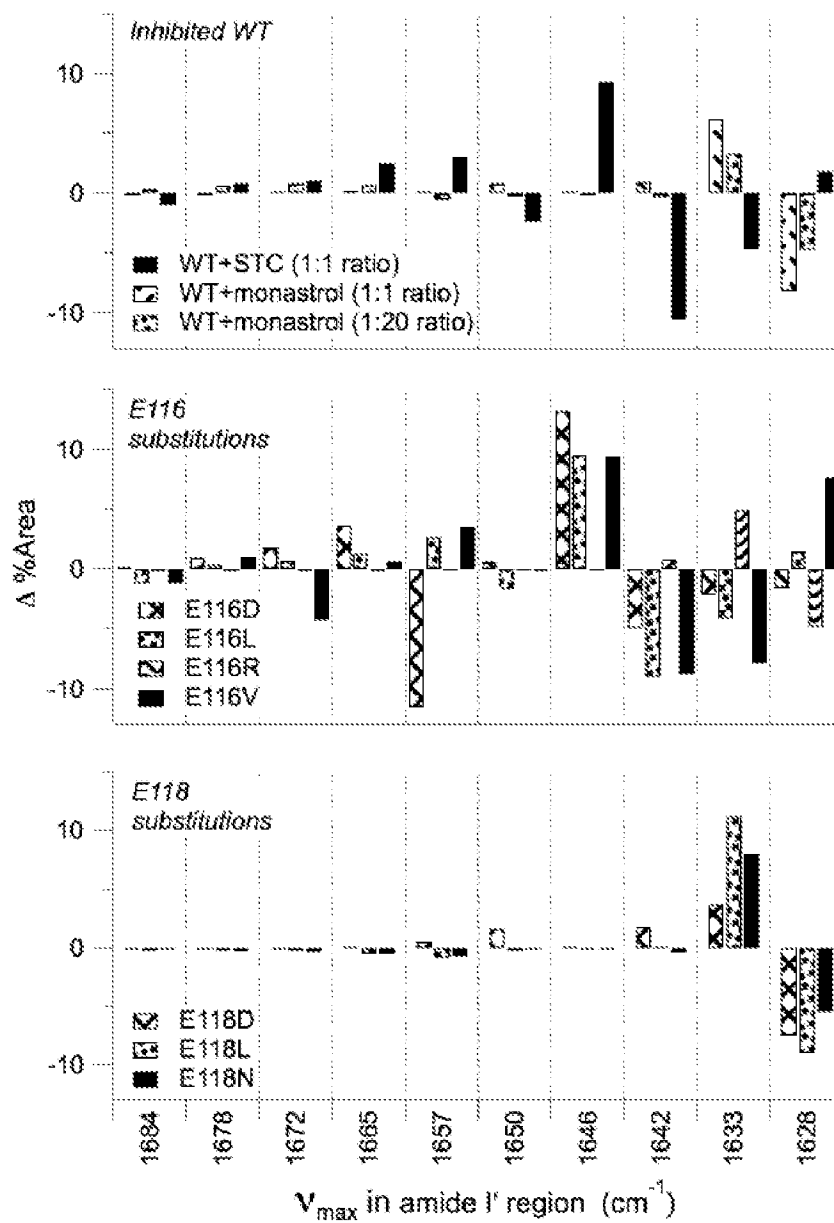
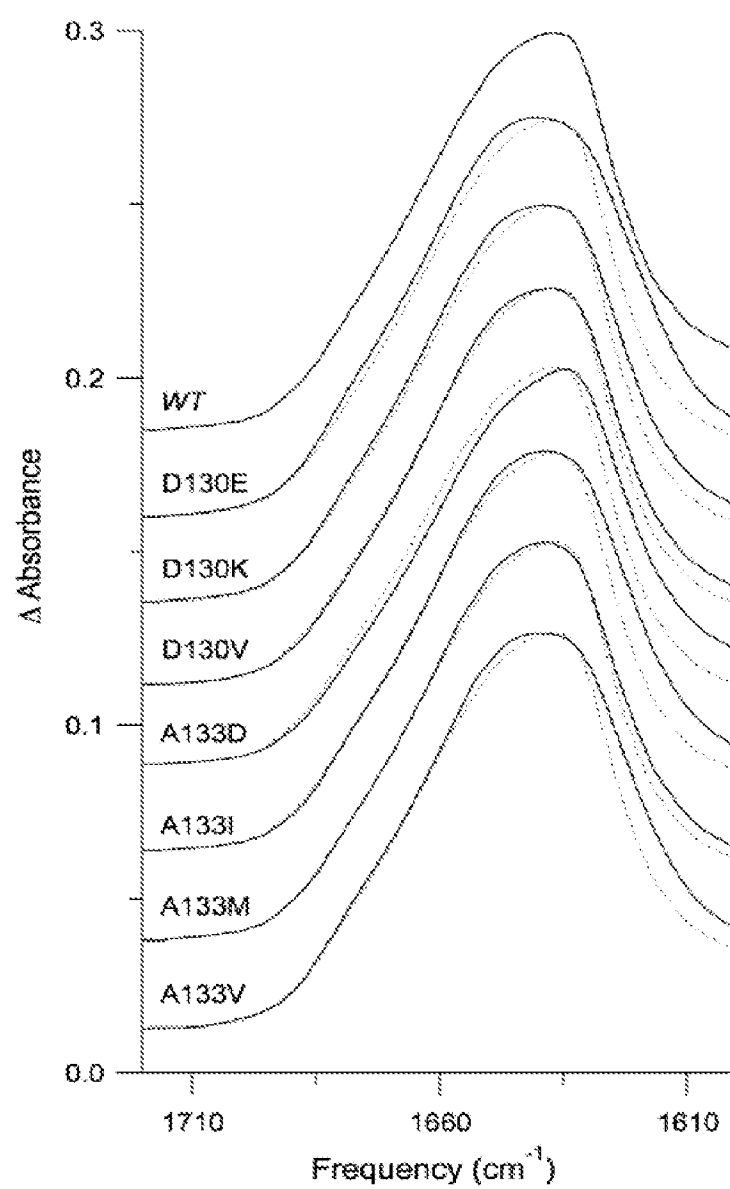


Fig. 4B

**Fig. 5A**

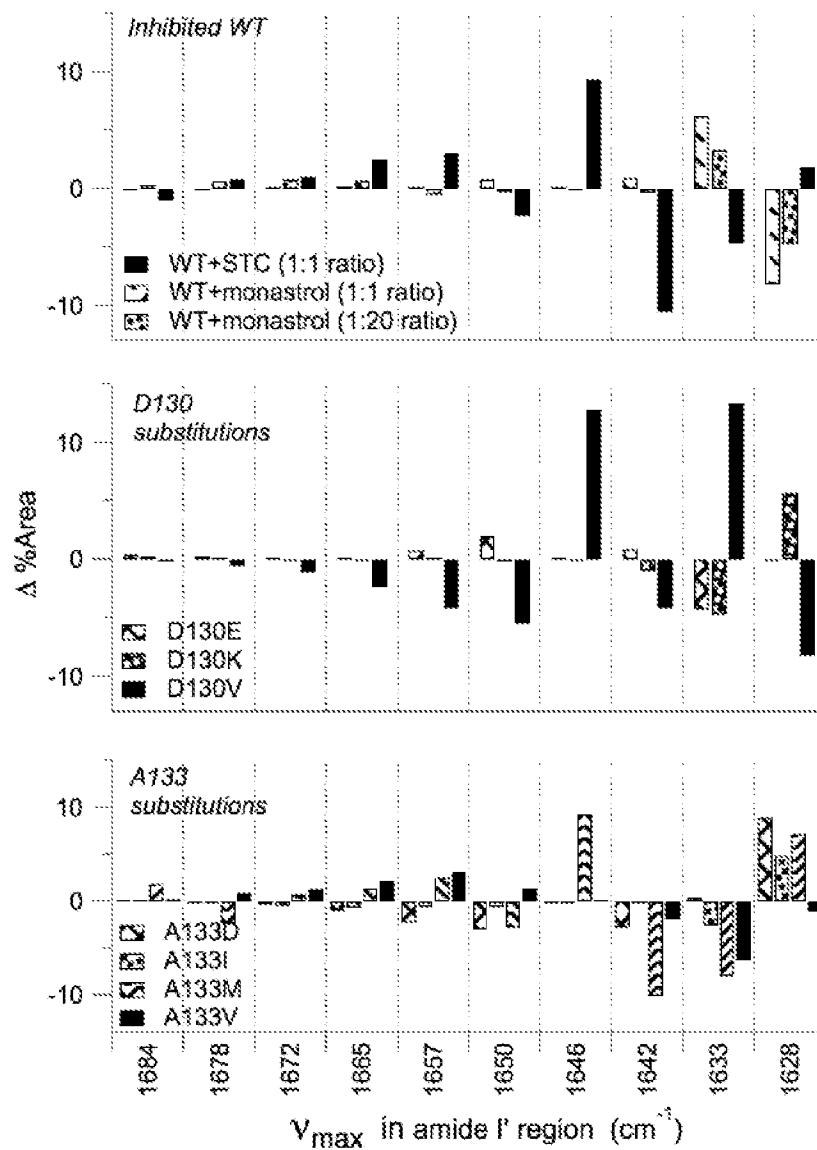
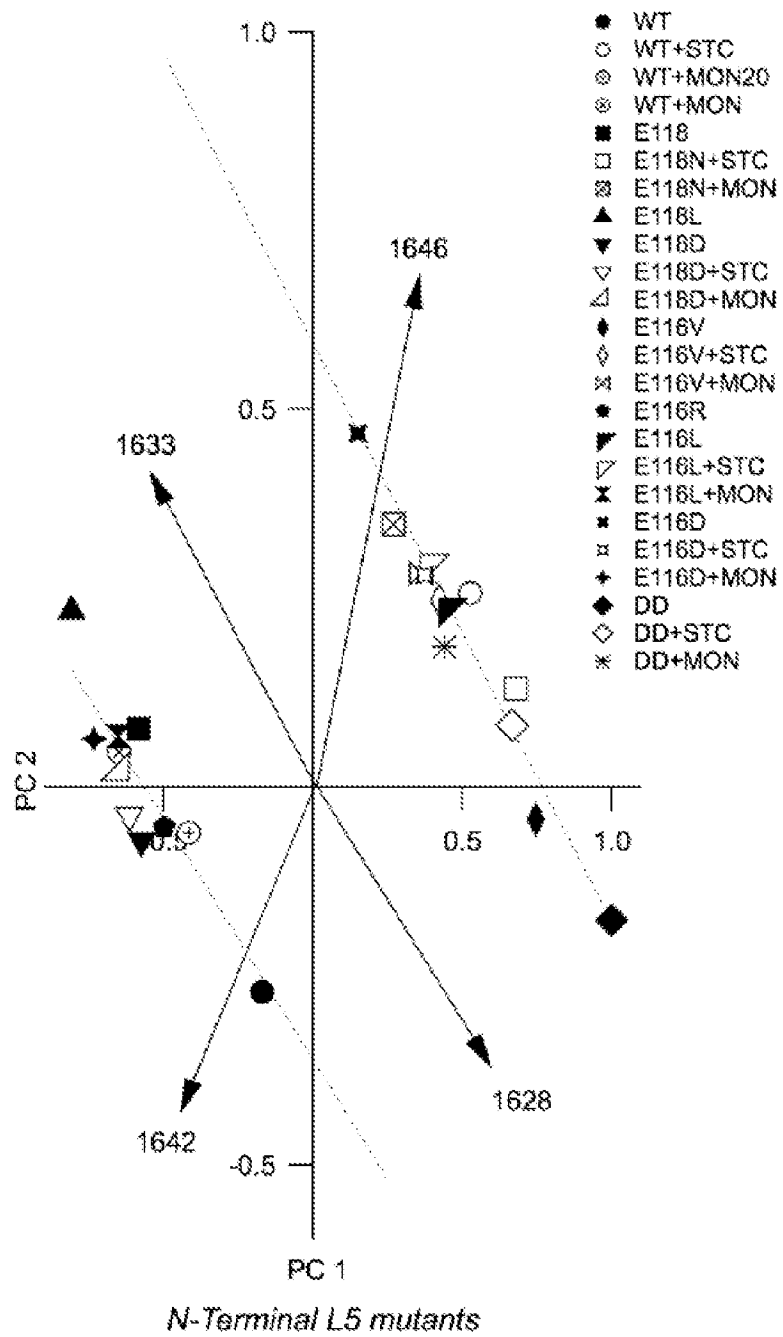


Fig. 5B

**Fig. 6A**

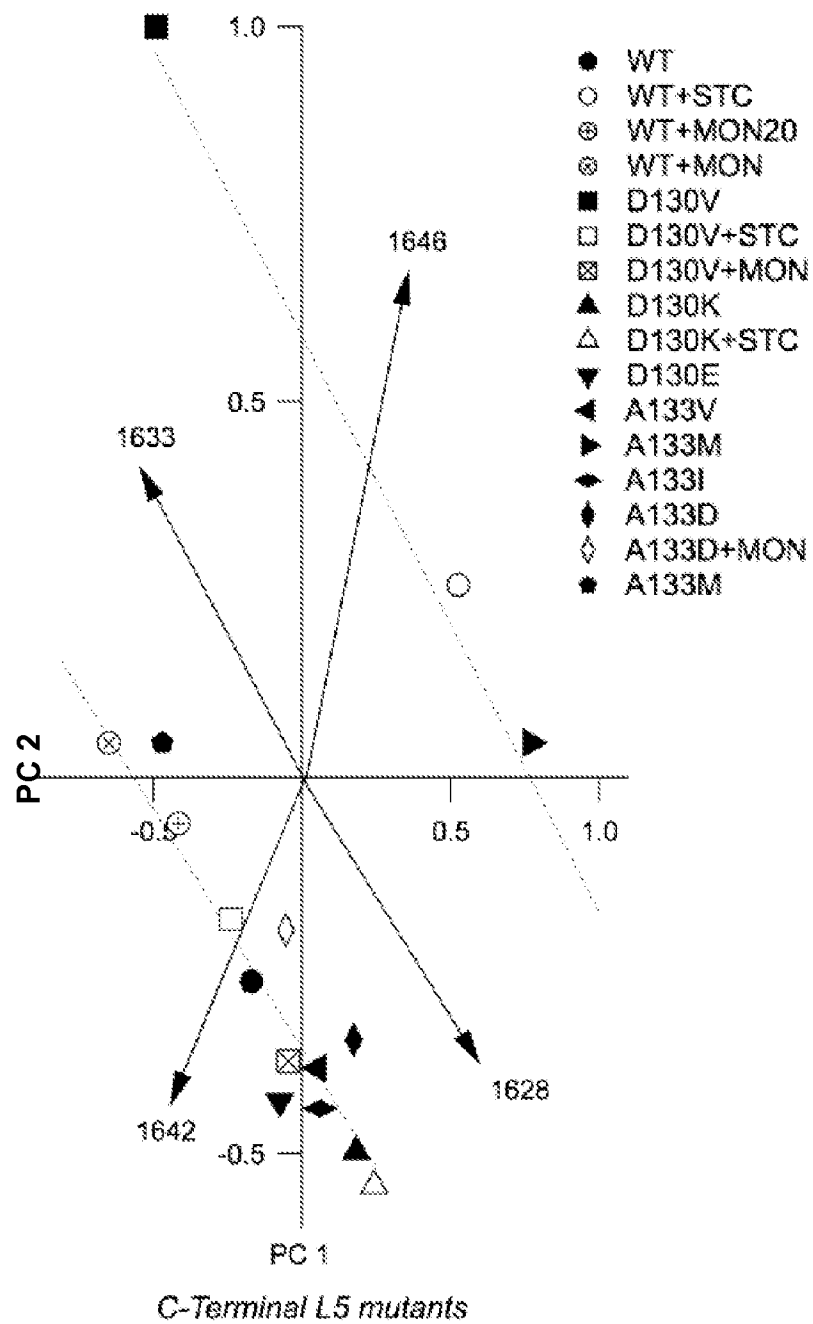


Fig. 6B

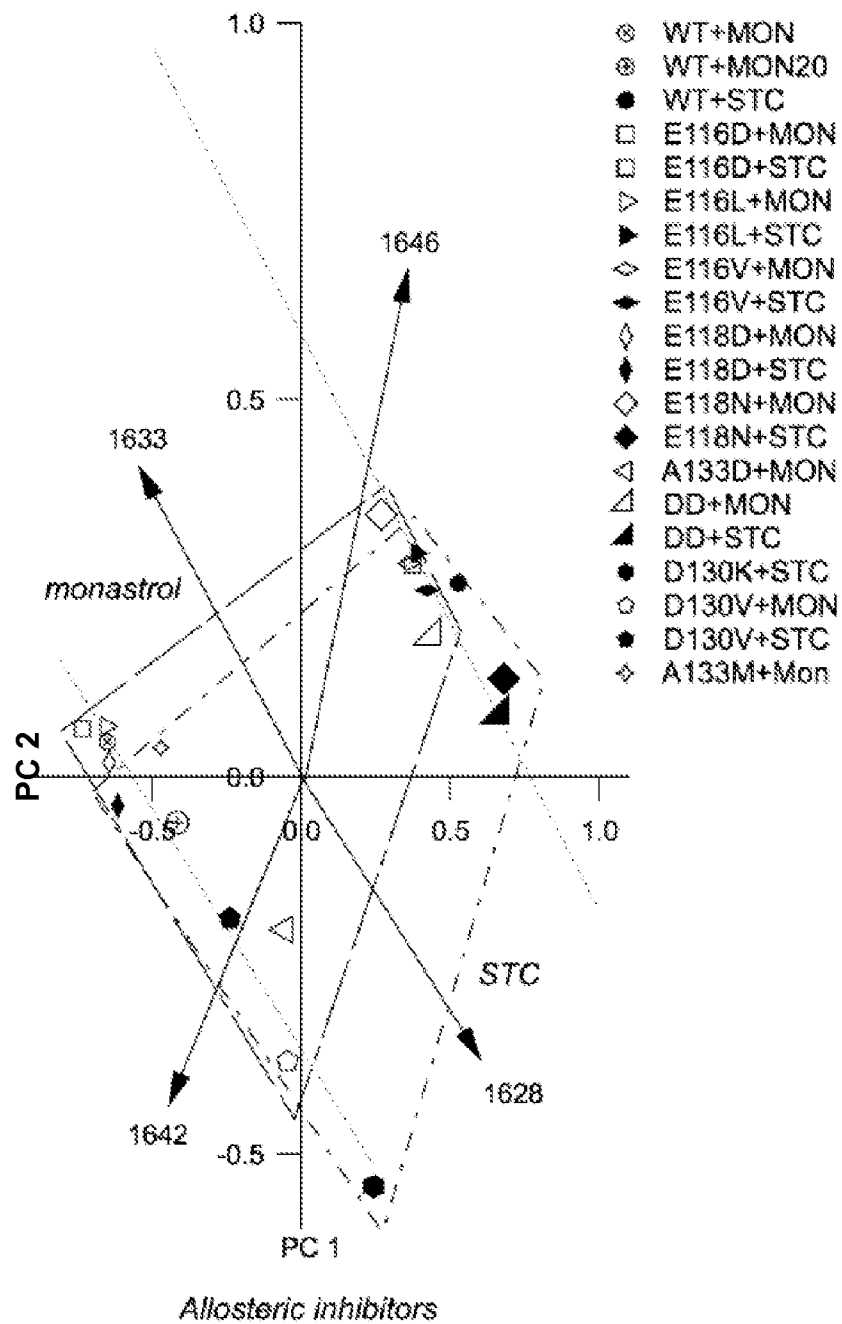


Fig. 6C