Identifying Molecular Mechanisms of Catechol O-Methyltransferase Activity and Regulation

Douglas Tsao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2011

Approved By:

Gary J. Pielak

Kevin M. Weeks

Alexandros G. Georgakilas

Luda Diatchenko

Nikolay V. Dokholyan

ABSTRACT

DOUGLAS TSAO: Identifying Molecular Mechanisms of Catechol O-Methyltransferase Activity and Regulation

Catechol O-methyltransferase (COMT) is a regioselective enzyme that functions by metabolizing catecholamines such as neurotransmitters and hormones *via* methylation at a single hydroxyl of the catechol moiety. Two cofactors are necessary in order to catalyze the reaction: S-adenosyl-L-methionine (SAM) and a divalent metal cation. Through computational docking studies, we show that a reason metal cations may be necessary for catalysis is due to its role in aligning the donating methyl group of SAM with the nucleophile of the substrate.

It is intriguing that an enzyme whose sole function is to degrade catecholamines contains highly conserved residues responsible for regioselective behavior at the catechol moiety, a function in which site-specific chemistry seems unnecessary. Our barrier height calculations for two different COMT ligands suggest that deprotonating the meta-hydroxyl leads to higher rates of methylation and consequently there is evolutionary pressure for selecting residues that can dock the ligand in an ideal conformation for efficient catalysis.

Our group previously identified three major haplotypes of COMT, where two single nucleotide polymorphisms (SNPs) produce synonymous changes and an additional SNP that creates a low enzymatic activity variant ($Val^{108}Met$). Here we show that the allelic variant encoding for the low activity protein shows the highest translational efficiency among the haplotypes, suggesting evolutionary selection of an RNA-structure destabilizing allele to compensate for the low activity mutation present within its protein structure. We provide a

mechanism whereby destabilizing alleles may facilitate translation initiation *via* computational modeling of each mRNA haplotype.

One of several biological factors that COMT influences is pain perception because of its role in degrading the neurotransmitters. Peripherally injected serotonin has been clinically shown to induce a hyperalgesic effect. Here we report that serotonin-induced hyperalgesia may be induced by inhibition of COMT. Our kinetic assays reveal serotonin as a non-competitive inhibitor with respect to catechol substrates. From computational modeling, we observe serotonin actively competing with the methyl donor S-adenosyl-L-methionine at the active site. Binding of COMT to serotonin inhibits the methyl donor from entering the active site, thus preventing methylation of COMT substrates.

ACKNOWLEGMENTS

I would like to thank the following people for their valuable support: my parents, for emphasizing the necessity of education as a means for success; my beautiful wife Melissa, for her unlimited support and patience throughout my graduate program; my advisors Nikolay Dokholyan and Luda Diatchenko, for their guidance and enabling me to work with a high level of scientific freedom; Feng Ding, for his extremely helpful discussions and technical expertise; and an additional acknowlegment to my Dad, for his enthusiasm of my research (he would probably read this dissertation in its entirety). Many thanks to my friends whom have provided me with moral support, and to my colleagues whom have aided me with my technical issues in the lab throughout the years.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ACRONYMS AND ABBREVIATIONS

5-HT	5 – Hydroxytryptamine
5-HT1A	5 – hydroxytryptamine receptor 1A
Å	Angstrom
A	Adenosine
AAAS	American Association for the Advancement of Science
AMBER	Assisted Model Building with Energy Refinement
APS	Average pain sensitivity
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
B3LYP	Becke 3-parameter exchange functional, Lee-Yang-Parr correlation
BCA	Bicinchoninic acid
°C	Degrees Centigrade/Celsius
C	Carbon
C	Cytosine
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary deoxyribonucleic acid
Cl	Chlorine
COMT	Catechol O-methyltransferase
DFT	Density functional theory
DMD	Discrete molecular dynamics
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
FMRP	Fragile X mental retardation protein
fs	Femtosecond
ΔG	Change in Gibbs free energy
G	Guanosine
h	Hours
H	Hydrogen
hCOMT	Human COMT
HPS	High pain sensitivity
K	Kelvin
k_B	Boltzmann constant
kcal	Kilocalorie
L	Liter
LPS	Low pain sensitivity
M	Molar
MB-COMT	Membrane-bound catechol O-methyltransferase

MD	Molecular dynamics
Mg	Magnesium
min	Minutes
mol	Mole
mRNA	Messenger ribonucleic acid
O	Oxygen
OH	Hydroxyl group
OR-486	3,5-dinitrocatechol
N	Nitrogen
Na	Sodium
NBO	Nonbonding orbital
nt	Nucleotide
pCMV	Porcine cytomegalovirus
PCR	Polymerase chain reaction
ps	Picosecond
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
rNTP	Ribonucleotide triphosphate
RPMI	Roswell Park Memorial Institute
S	Sulfur
SAM	S-adenosyl-L-methionine
S-COMT	Soluble catechol O-methyltransferase
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SHAKE	Algorithm for adjusting atomic positions to satisfy constraint conditions in molecular dynamics - J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, J. Comput. Phys., 23, 327 (1977)
S_N2	Bimolecular nucleophilic substitution
SNP	Single nucleotide polymorphism
SPR	Surface plasmon resonance
SSRI	Selective serotonin reuptake inhibitor
T	Thymine
TIP3P	Transferable intermolecular potential three-point
T_m	Melting temperature
TMJD	Temporomandibular joint disorder
TS	Transition state
U	Uracil
UTR	Untranslated region
vdW	van der Waals
WHAM	Weighted histogram analysis method

CHAPTER I: GENETIC AND STRUCTURAL FEATURES OF CATECHOL O-METHYLTRANSFERASE

Structural Characterization of Catechol O-Methyltransferase

Catechol O-methyltransferase (COMT) is an enzyme found in virtually all mammalian tissues and is responsible for metabolizing catechols (either endogenous or exogenous in origin) by methylating a single hydroxyl group at either the meta- or para- position (1). Two additional cofactors are required for the reaction to occur: S-adenosyl-L-methionine (SAM) and a divalent metal cation. Magnesium serves as the metal catalyst within the cell, although *in vitro* studies have shown that other metals are sufficient for activity (2). Methylation of catechols proceeds via a S_N2 mechanism, where the catechol is first deprotonated at a single hydroxyl and the resultant oxyanion attacks the cationic methyl group thereby displacing it from the SAM cofactor (3).

There are two main isoforms of COMT: a soluble (S-COMT) and membrane-bound (MB-COMT) isoform (1). The predominant isoform within the human body is S-COMT. The MB-COMT isoform is mostly present at the human brain, and is bound to the membrane of the endoplasmic reticulum as opposed to the cellular membrane (4,5). With the exception of the first 50 residues, the two sequences remain identical since they derive from the same gene. Two distinct promoters are responsible for the isoforms, where transcribing the first promoter at the third exon results in the MB-COMT isoform, while the distal promoter for S-COMT is approximately 200 nucleotides downstream on the third exon (6). From secondary structure predictions, it is believed that the first 50 residues of MB-COMT form a transmembrane α-helix (7).

The crystal structure of S-COMT (and henceforth COMT) was originally solved in 1994 for rats (PDB: 1VID) (8), with a human isoform eventually solved in 2008 (PDB: 3BWM) (9). The topology of COMT resembles a Rossman fold, with a 7-stranded mostly parallel β-sheet sandwiched between 5 α-helices at the N-terminal side and 3 α-helices at the opposing side (Figure 1.1A). For COMT, the Rossman fold topology is used for recognition and binding of the nucleotide cofactor SAM. Apo-COMT initially binds to SAM, as its binding site is recessed deepest inside the active site.

Adjacent to the SAM binding site is a group of negatively charged residues (D141, D169, E199) and the carbonyl of N170 that coordinate to magnesium (Figure 1.1B). The exact coordination of Mg^{2+} within the active site is a topic of debate in the field (10,11). From the crystal structure, two coordination sites are occupied by each hydroxyl from the catechol. However, an inhibitor is used to prevent methylation of the substrate for all crystal structures of COMT. These inhibitors act to prevent methylation by reducing the nucleophilicity of the oxyanion that attacks the methyl group of SAM. For many of the inhibitors (including those used for crystallography), nitrates are added to the catechol ring for this purpose. These nitrate groups reduce the nucleophilicity due to the number of resonance structures formed once a hydroxyl is deprotonated. Therefore, it is thought that the catechol coordinates twice to magnesium because the nucleophilicity of the catechol is dampened. Several groups have demonstrated that molecular dynamics (MD) simulations of a natural catechol substrate with COMT result in a complex where magnesium is mono-coordinated with the catechol (10,11). The exact hydroxyl that is coordinated with magnesium is a further topic of debate, where one group obtains a complex with the oxyanion coordinated to magnesium (11) and another group obtains a complex with the neighboring hydroxyl group coordinated to magnesium (10).

Nonetheless, Mg^{2+} remains hexacoordinated next to the sulfonium center of SAM where the last coordination site is occupied by water.

The catechol substrate is the third and final ligand bound to COMT prior to methylation (12). Its binding site is therefore the most solvent accessible of the three ligands (9). The β-turn separating anti-parallel strands 6 and 7 form a loop that recognizes catechol substrates. Residue L198 within this catechol-binding loop make direct contact with the catechol ring itself. Further along the loop, residues R201 and E202 form the outer pocket. A loop that connects β-strand 6 and α-helix 8 contains a proline residue at position 174 that stacks with the catechol ring in an edge-to-face interaction.

A region of the protein known as the "hydrophobic wall" consists of residues that bind to the side-chain of a catechol substrate (8). Residues E199 and Y200 of the catechol-binding loop have been found in MD simulations to migrate toward the catechol side-chain (11). Two additional residues outside the two aforementioned loop domains form van der Waals (vdW) contacts with the catechol: M40 and W38, located on the loop between α-helices 2 and 3. Tryptophan 143, residing at the tail end of α-helix 7, has been shown via MD simulations to bind the side-chain of catechols, potentially via a cation-π interaction (11). These residues of the hydrophobic wall are important not only for binding to the substrate, but they are responsible for the regioselective properties of the enzyme (11). Specifically, interactions of the side-chain with residues of the hydrophobic wall guide the substrate into the active site. Therefore, the catechol is oriented into the active site in such a way that the side-chain interactions are most favorable. For many biological ligands, the side-chain of catechol substrates are closer to the 3'OH. Consequently, many COMT products are methylated at the meta-position.

A common variant of COMT has a mutation at position 108 (158 for MB-COMT) where valine is substituted for methionine (1,13,14). Interestingly, this residue is located over 15 Å away from the active site and is exposed on the solvent-accessible surface of the protein (9). Yet it has been shown from several studies that the ^{108}Met variant displays lower enzymatic activity within the cell (1,12,15). The protein destabilization resulting from the substitution varies in species; there is approximately a 4 kcal/mol destabilization for rats, and only a 1 kcal/mol destabilization for humans (16). Furthermore, a comparison of the crystal structures between ^{108}Val and ^{108}Met reveals a root mean square deviation (RMSD) of only 0.2 Å (9). However, MD simulations performed for each variant reveal some differences in their dynamics (16). The ^{108}Met variant exposes more residues to the solvent, thereby potentially facilitating its degradation through easier access to proteases.

Three Main Haplotypes of COMT

The Val^{108}Met polymorphism is not the exclusive variant known for COMT. Overall, there are three known haplotypes that were shown to have functional effects and encompass 96% of the human population. These haplotypes were first discovered in a pain-association study, and therefore are named according to their originally associated phenotypes (Figure 1.2A): low pain sensitivity (LPS), average pain sensitivity (APS), and high pain sensitivity (HPS) (13). The most common in Caucasian population is APS haplotype that encodes for the ^{108}Met variant that has lower enzymatic activity. The next frequent haplotype is LPS, which corresponds to the wild-type COMT protein. HPS exhibits the lowest activity of the three haplotypes; however, there is no difference in the protein structure compared to the wild type. Only a single synonymous mutation (rs4818) differentiates LPS and HPS within the coding region of the gene.

The low enzymatic activity displayed by HPS haplotypes is a direct result of its low protein abundance (Figure 1.2C,D) (15). Secondary structural predictions of the mRNA transcript suggest a highly stable stem-loop structure formed exclusively by the HPS haplotype (Figure 1.2B) (15). These predictions were confirmed through further mutagenesis studies where point mutations were made along the LPS transcript to induce the HPS stem-loop structure, and rescue mutations were performed for the HPS haplotype to recapitulate the LPS stem-loop structure. Mutants that were predicted to have the HPS stem structure showed low protein expression, while the rescue mutants that were predicted to have the LPS stem structure displayed higher protein expression. Therefore, it was concluded that HPS haplotypes exhibit low protein abundance in the cell due to a highly stable stem-loop on its mRNA transcript. A stem-loop that possesses a melting temperature (T_m) higher than 70 °C can inhibit its translation due to the high-energy requirements needed for the ribosome to unwind this region.

Biological Significance of COMT Haplotypes

Those with the wild-type COMT haplotype experience the least amount of pain sensitivity due to the enzyme's higher activity (13,15). Neurotransmitters such as dopamine, epinephrine, and norepineprine are endogenous substrates of COMT due to their catecholamine structure (14). Sustained levels of catecholamines are known to induce a pain response by binding to a class of G-protein coupled receptors known as β-adrenergic receptors. Specifically, it has been shown that β_2- and β_3-receptors are responsible for the low COMT activity dependent enhanced pain sensitivity (17). β-adrenergic receptors are activated by catecholamines and initiate the signaling cascade by stimulating cAMP production. When methylated by COMT,

catecholamines lose their biological activity, and β_2- and β_3-receptors activities decrease. The decrease in β_2 and β_3 receptor activity is associated with a lower perception of pain.

The three haplotypes of COMT have been associated chronic pain phenotypes including temporomandibular joint disorder and fibromyalgia (13,18). Furthermore, the probability of post-operative pain increases when patients possess alleles for lower COMT activity (19). However, other factors are associated with COMT including the development of white brain matter and drug response (20,21). For instance, a common treatment for the symptoms of Parkinson's disease is the prescription of levodopa. Levodopa contains a catecholamine structure that undergoes methylation by COMT. Treatment of patients with COMT haplotypes that have a higher enzymatic activity is less effective unless a COMT inhibitor is co-administered.

Summary

Studying structural, molecular, and genetic mechanisms of COMT activity and regulation are thus important elements of understanding human pain perception and its variability within the human population. Moreover, the biological significance of COMT spans beyond pain perception due to the fundamental role that COMT plays in degrading both endogenous and exogenous catechols, thereby making it critical for many other phenotypes including blood pressure, moods and drug dependence. Effective diagnosis and treatments of individuals will improve as our understanding of COMT advances.

In the next two chapters, we further explore the structural details of COMT activity. The mechanism of how SAM is initially accommodated within the COMT active site is discussed, followed by a detailed study of the regioselectivity of COMT. Next, we examine how COMT is

regulated at the mRNA level and in the presence of serotonin in chapters 4 and 5, respectively. Our work on COMT regioselectivity and mRNA regulation have been recently published (22,23), while the manuscripts for SAM binding and serotonin regulation are currently under preparation.

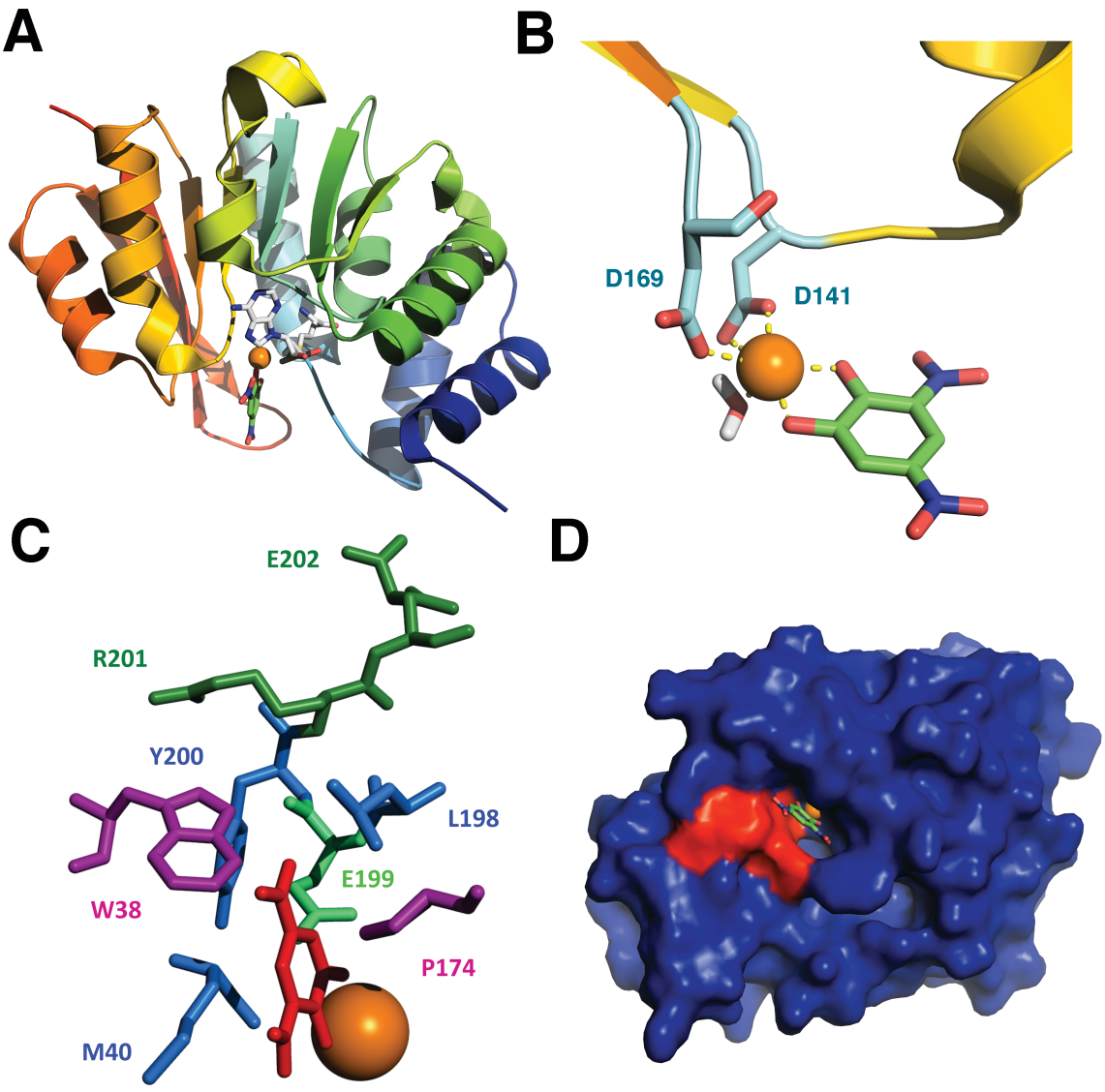


Figure 1.1. Structural features of S-COMT. (**A**) Overall topology of S-COMT as represented in cartoon format. Protein is colored from blue (N-terminus) to red (C-terminus). 3,5-dinitrocatechol is colored green, Mg^{2+} is colored orange, and SAM is colored gray. (**B**) Magnesium coordination within COMT active site (as depicted from the crystal structure). Coordination bonds are represented by dashed yellow lines. (**C**) Interactions between 3,5-dinitrocatechol and COMT residues. 3,5-dinitrocatechol is shown as red. Blue residues define the inner pocket, green residues define the outer lip, and purple residues are classified as the

"gatekeeper" residues, as defined by Rutherford et al (9). (**D**) Surface representation of 3,5-dinitrocatechol bound to COMT. The "hydrophobic wall" residues are highlighted in red.

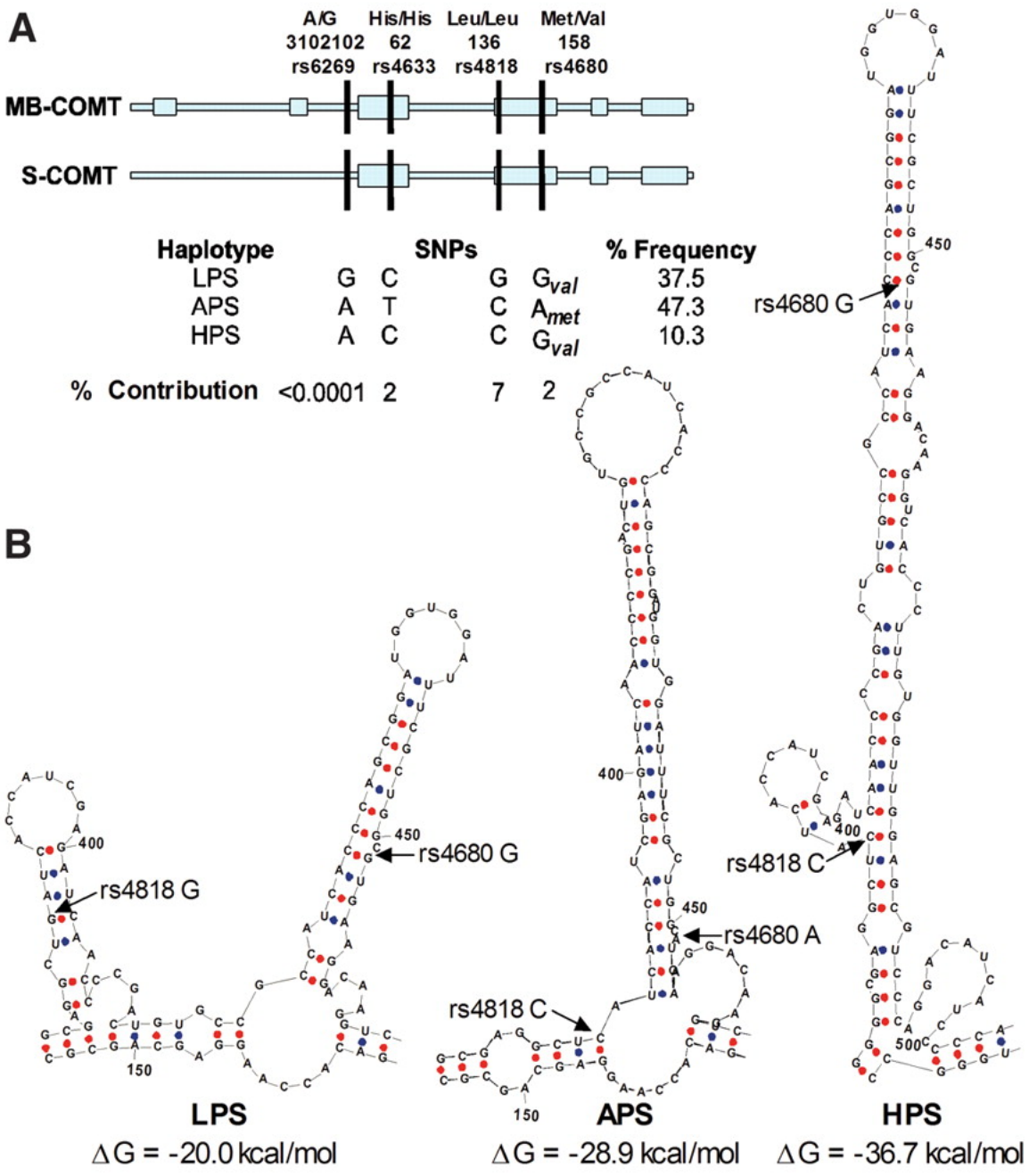

A
A/G
3102102
rs6269
His/His
62
rs4633
Leu/Leu
136
rs4818
Met/Val
158
rs4680
MB-COMT
S-COMT
Haplotype
SNPs
% Frequency
LPS G C G G_{val} 37.5
APS A T C A_{met} 47.3
HPS A C C G_{val} 10.3
% Contribution <0.0001 2 7 2
B
rs4818 G
rs4680 G
rs4818 C
rs4680 A
rs4818 C
rs4680 G
LPS
ΔG = -20.0 kcal/mol
APS
ΔG = -28.9 kcal/mol
HPS
ΔG = -36.7 kcal/mol
C
S-COMT Activity
(relative pg/mlNMN protein)
30
20
10
0
+++
***
***
LPS
APS
HPS
D
S-COMT
β-actin
LPS
APS
HPS
Cntl

Figure 1.2. Haplotypes of COMT. From "Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*, **314**, 1930-1933". Reprinted with permission from AAAS. (**A**) Diagram of MB-COMT and S-COMT gene with common SNP locations for each haplotype. Frequencies were derived from a cohort study of 202 healthy Caucasian females. Large rectangles along the gene represent exon regions while the smaller regions are intronic. SNPs are highlighted by a vertical black line. (**B**) Predicted stem-loop structure for COMT haplotypes. Location of SNPs are highlighted with the corresponding arrows. (**C**) Enzymatic activity of each haplotype within a PC-12 (rat pheochromacytoma) cell lysate. (**D**) Expression levels of each haplotype within PC-12 cells.

CHAPTER II: STRUCTURAL MECHANISM OF S-ADENOSYL METHIONINE BINDING TO CATECHOL O-METHYLTRANSFERASE

Douglas Tsao,[1,2] Luda Diatchenko,[2] and Nikolay V. Dokholyan[2,3]*

[1]*Department of Chemistry*

University of North Carolina, Chapel Hill, NC 27599, USA;

[2]*Center for Neurosensory Disorders, School of Dentistry*

University of North Carolina, Chapel Hill, NC 27599, USA;

[3]*Department of Biochemistry and Biophysics, School of Medicine*

University of North Carolina, Chapel Hill, NC 27599, USA

*Corresponding Author: Nikolay V. Dokholyan (dokh@med.unc.edu)

Keywords: COMT, SAM, methyltransferase, molecular docking, alternative conformation

ABSTRACT

Methyltransferases possess a homologous domain that requires both a divalent metal cation and S-adenosyl-L-methionine (SAM) to catalyze its reactions. The kinetics of several methyltransferases has been well characterized; however, the details regarding their structural mechanisms have remained unclear to date. Using catechol O-methyltransferase as a model, we perform computational docking simulations to elucidate the initial stages of cofactor binding. We find that SAM adopts a different docking pose in

the absence of metal and substrate in comparison to the holoenzyme, illustrating that the metal reorients SAM in a conformation suitable for donating its methyl substituent to the recipient ligand.

INTRODUCTION

Catechol O-methyltransferase (COMT) is a metalloenzyme that metabolizes biologically active catechol-containing structures by methylation of a single hydroxyl group (1). A number of neurotransmitters contain a catecholamine moiety and are deactivated by COMT. Activity of COMT is thus correlated with many critical biological functions including cognition (24), stress response (25), and pain sensitivity (13).

Crystallographic and enzymatic studies of COMT have given significant insight to its mode of activity (9,12). The cofactor S-adenosyl-L-methionine (SAM), responsible for donating the methyl group to the catechol, initially binds to the enzyme. This complex then binds a divalent metal cation through coordination bonds to several acidic residues and a single water molecule in the active site. Additional residues then bind to the catechol substrate, and the metal coordinates to the hydroxyl group inside the active site. Although the general kinetic mechanism for COMT substrate binding is known, the structural details that govern binding have yet to be identified. Current crystal structures of COMT are bound to SAM, Mg^{2+}, and an inhibitor. Using computational docking tools (Methods; (26,27)), we dock SAM to an apo-COMT protein to determine the initial binding poses. We also perform docking simulations with SAM inside a COMT•metal•catechol complex (holo-COMT) and compare to the crystal structure. From our results, we show that the metal cation is critical for aligning SAM in a conformation suitable for

catalysis with the substrate. Here we detail a mechanism whereby SAM initially binds to COMT, and upon metal binding, subsequently reorients itself for the enzyme to accommodate the substrate and facilitate methyl transfer.

METHODS

We employ the MedusaDock package to generate possible ligand conformations within the protein active site, and utilize the MedusaScore package to evaluate each conformation generated by MedusaDock (26,27). MedusaDock enables flexible docking of both the ligand and the side-chain amino acids of the protein. We perform all ligand docking simulations using the crystal structure of human COMT (PDB: 3BWM) (9).

Prior to docking S-adenosyl methionine (SAM) to apo-COMT, we stripped all crystallized ligands bound to COMT. Crystallographic waters that were in the active site were retained for our docking simulations. We performed 200 docking simulations, where each simulation began with a different seed number and each conformation generated from a simulation was subsequently minimized. All conformations were then ranked according to their free energy value. The lowest energy structure is determined to be the native pose.

To model the entire complex with SAM, Mg^{2+}, and catechol bound (holo-COMT), we add several constraints and make several modifications. First, we replace Mg^{2+} with Zn^{2+} since there are no Mg^{2+} parameters previously defined in the MedusaScore force field. However, previous experiments suggest that Zn^{2+} is a suitable alternative for Mg^{2+} as it is ~80% as effective (2). The general mechanism is that the metal divalent ion displaces the monovalent cationic amine group and acts as a steric block for the rest of the methionine side-chain. We place constraints to fix the position of the metal and catechol as found in the crystal structure,

while we flexibly dock SAM into the active site. We find that the lowest energy pose recapitulates the crystal structure, with a heavy atom root mean square deviation of 0.68 Å.

RESULTS AND DISCUSSION

In the absence of metal and substrate, portions of SAM bind to a different groove of COMT (Figure 2.1A). The adenosine moiety of SAM remains identical to the holo-COMT structure, with the Ile91 side-chain packing on top of the pyrimidine portion of adenine and the imidazole of His142 participating in a perpendicular edge-to-face aromatic interaction with the pyrimidine (Figure 2.1B). The imidazole of adenine participates in an additional edge-to-face interaction with the indole side-chain of Trp143. Additional polar contacts are made between residues 118-120 and the purine nitrogens. The two hydroxyl groups of the ribose participate in hydrogen bonding interactions with the side-chain of Glu90.

Similarities between the structures of SAM docked inside apo-COMT and holo-COMT end at the sulfonium center. In the apo-COMT complex, the terminal amine group of SAM is involved in a hydrogen-bonding network with the side-chains of Asp141 and Asn170 and the backbone carbonyl of Met40 (Figure 2.1C). The terminal carboxyl group participates in hydrogen bonding with Lys144 and a single water molecule. The interactions highlighted here with methionine underlie an important point for why catalysis cannot occur without a metal. Lys144 is responsible for deprotonating the catechol to create the oxyanion responsible for attacking the methyl group. In this particular pose, the Lys144 is preoccupied in a hydrogen-bonding network and is unavailable to deprotonate.

Prior to catechol binding, a divalent metal cation must first displace the positively charged amine group. Several divalent cations are capable of contributing to enzymatic activity,

although the native metal is magnesium in vivo (2). Here we have modeled the divalent metal cation using Zn^{2+}, which is 80% as effective as magnesium. We find that upon metal binding, steric occlusion prevents the methionine of SAM from binding to the negatively charged pocket and is forced into an interior groove. The amine group of the methionine maintains a hydrogen bond with Asp141, albeit at a different position, but additionally forms hydrogen bonds with the backbone carbonyl of Gly66 and the side-chain of Ser72. The amide backbones of Ser72, Val42, and an additional water molecule form hydrogen bonds with the carboxyl group of SAM. Additional hydrophobic interactions are formed between the methionine side-chain and residues 40, 42, 60, 68, and 89 inside this pocket (Figure 2.1D).

The results found here are initially surprising because it is expected that SAM binds to apo-COMT as found in the crystal structure of the holoenzyme complex. The amine and carboxyl tail of SAM form favourable van der Waals contacts and satisfy their hydrogen bonds in both conformations. However, the carboxylate side-chains within the active site preferentially bind to the amine side-chain of SAM, as the polar contacts are stronger than those shared with the backbone carbonyl groups of holo-COMT.

The role that magnesium plays in COMT activity has remained unclear to date. Here we show that a divalent metal cation is essential for structural rearrangements of the SAM cofactor. Without any metal bound to COMT, methylation is 6.2% as effective as compared to with all cofactors present (2). The orientation of the methionine without metal present positions the donating methyl group 6.1 Å away from the oxyanion, compared to a separation of 2.7 Å in the presence of metal (Figure 2.1E). Furthermore, the oxyanion and sulfonium no longer form a 180° angle in between the methyl group. This angle decreases to ~45° in the absence of metal.

Therefore binding to SAM and substrate alone produces no methylated product in the active site of COMT.

The mechanism described here is of broad biological interest since SAM serves as a common methyl donor for many methylation reactions, including CpG methylation (28). The binding site of SAM on COMT is homologous with many SAM-dependent methyltransferase structures, with several unique residues that form part of the catechol-binding site. Thus our mechanism could be applicable to a broad range of methyltransferases that require a divalent metal cation and SAM.

ACKNOWLEDGMENTS

This work was supported by the US National Institutes of Health grant [R01GM080742 to N.V.D.]; American Recovery and Reinvestment Act supplements [GM080742-03S1, GM066940-06S1 to N.V.D.]; National Institute of Dental and Craniofacial Research and National Institute of Neurological Disorders and Stroke grants [RO1-DE16558, UO1-DE017018, PO1 NS045685 to L.D.].

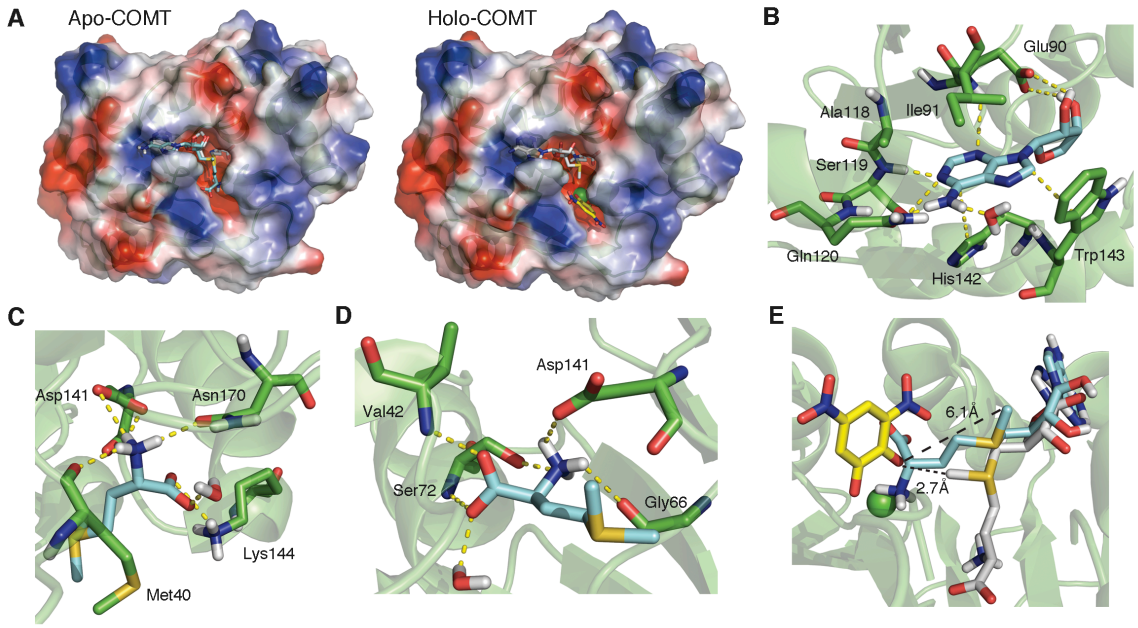


FIG. 2.1. Binding configurations of SAM with and without metal. (A) Alignment of SAM poses with and without metal. Cyan structure represents without metal (apo-COMT) and gray structure represents SAM in the presence of metal. In the right panel, the yellow structure is dinitrocatechol and green sphere is Zn^{2+}. Surface of protein is shown as an electrostatic map with blue regions representing clusters of positive charge and red representing regions of negative charge. (B) Contacts made with adenosine portion of SAM (identical for apo- and holo-COMT). Cyan structure is adenosine and green structures are corresponding residues. Yellow dashed lines indicate contacts. (C) Contacts made with methionine portion of SAM (apo-COMT). Cyan structure is methionine portion of SAM. (D) Contacts made with methionine portion of SAM (holo-COMT). (E) Comparison of angle between methyl donor and accepting hydroxyl in the presence and absence of metal. Cyan structure is apo-COMT and gray structure is holo-COMT.

CHAPTER III: REGIOSELECTIVITY OF CATECHOL O-METHYLTRANSFERASE CONFERS ENHANCEMENT OF CATALYTIC ACTIVITY

Douglas Tsao[a], Shubin Liu[b], Nikolay V. Dokholyan[c]*,

[a] Department of Chemistry,

University of North Carolina, Chapel Hill, North Carolina 27599;

[b] Research Computing Center,

University of North Carolina, Chapel Hill, North Carolina 27599;

[c] Department of Biochemistry and Biophysics,

University of North Carolina, Chapel Hill, North Carolina 27599

Corresponding Author: Nikolay V. Dokholyan (dokh@med.unc.edu)

Abstract

Catechol O-methyltransferase (COMT) metabolizes catechol moieties by methylating a single hydroxyl group at the meta- or para- hydroxyl position. Hydrophobic amino acids near the active site of COMT influence the regioselectivity of this reaction. Our sequence analysis highlights their importance by showing that these residues are highly conserved throughout evolution. Reaction barriers calculated in the gas phase reveal a lower barrier during methylation at the meta- position, suggesting that the observed meta-regioselectivity of COMT

can be attributed to the substrate itself, and that COMT has evolved residues to orient the substrate in a manner that increases the rate of catalysis.

1. Introduction

Catechol O-methyltransferase (COMT) is an enzyme that is primarily responsible for the deactivation of biologically active and toxic catechols and is found in virtually all mammalian tissues (1). Many neurotransmitters and hormones within the body are characterized by a catechol or catecholamine structure and hence serve as substrates for COMT. Deactivation of substrates proceeds by methylation of a single hydroxyl to form the meta- or para- product (29).

The active site consists of the catechol ligand, which is mono-coordinated to Mg^{2+}, and the co-enzyme S-adenosyl methionine (SAM) that serves as the methyl donor. Conversion to the guaiacol product proceeds via an S_N2 mechanism, where Lys144 serves as a catalytic base and deprotonates a single hydroxyl to form the oxyanion that attacks a methyl group from the sulfonium. Several groups have calculated the reaction barrier for catechol to be approximately 20 kcal/mol in height (10,11,30).

Meta-methylation is the preferential product of COMT with the extent of regioselectivity dependent upon the catechol side-chain (Table 3.1) (12). There has been no reported ligand where para-methylation is favored. It has been largely believed that the preference for meta-methylation stems from side-chain interactions with the residues comprising a hydrophobic wall near the active site. Previous molecular dynamics simulations have proposed that Trp143 forms a cation-π interaction with the substrate, conferring additional stability to the meta- position

exclusively (11). Migration of the Trp143 residue from binding to the adenosine motif of the SAM cofactor (as found in the crystal structure) to binding the side-chain of the catechol ligand demonstrates how the active site dynamics of COMT can modulate regioselectivity (31).

It is intriguing that an enzyme whose sole function is to degrade neurotransmitters and hormones displays specific regioselective behavior at the catechol moiety, a function in which site-specific chemistry seems unnecessary. Additionally, it has been assumed throughout the literature that regioselectivity has been imposed by the enzyme alone and that the ligands themselves have no preference (1,9,12). Here, we examine to see if the ligands possess intrinsic regioselective properties and whether the protein has adapted to take advantage of such properties by using regioselection as a method to enhance the rate of catalysis. Three models are constructed for this study: Model 1) a simple model composed of the catecholamine and trimethylsulfonium; Model 2) a model of catecholamine and trimethylsulfonium with four solvation shells of water; and Model 3) a model of the protein's active site composed of the residues Trp143 and Lys144, Mg^{2+} coordinated to crystallographic water and surrounding residues, and the catecholamine and trimethylsulfonium. We compute the reaction barriers for methylating dopamine and levodopa at the meta- and para- positions using DFT since they represent two extremes of regioselection with meta/para ratios of approximately 4 and 16, respectively.

2. Methods

2.1. Sequence conservation of COMT

Sequence alignments of COMT were performed for 13 different species: *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Equus caballus*, *Sus scrofa*, *Rattus norvegicus*, *Mycobacterium vanbaalenii*, *Pichia stipitis*, *Mycobacterium gilvum*, *Nicotiana tabacum*, *Salmo salar*, *Populus trichocarpa*, *Papaver somniferum*, *Thalictrum tuberosum*. The software ClustalX 2.0.10 was used to perform the alignment using the Neighbor Joining algorithm with default parameters.

2.2. MD simulation of ligand complexes

The crystal structure of human soluble-COMT (PDB: 3BWM) was used as a starting point for our MD simulations. A caveat of the crystal structure, as published by several groups, is that the inhibitor used to crystallize the protein has a different binding geometry compared to natural ligands. Specifically, since most inhibitors decrease the nucleophilicity of the attacking oxygen with nitrate groups, these ligands bind bidentate to the Mg^{2+} in the active site. Several groups have proposed that natural substrates of COMT contain a monodentate coordination number in the active site (10,11). Additionally, the interaction between Trp143 and the ligand side chain is possible only after MD simulations (11).

Thus, we initially dock either dopamine or levodopa into the active site using the 3,5-dinitrocatechol substrate position as a reference. We adopt a similar simulation methodology as outlined by Kuhn et al in order to obtain our minimized complexes (11). Briefly, we simulate

our system using the AMBER force field (amber03) and derived constraints for our ligands and S-adenosylmethionine using Antechamber (32). Parameters for Mg^{2+} were adjusted according to Kuhn's values; i.e., an atomic charge of +2, R^*_{Mg} = 0.7868 Å, ε_{Mg} = 0.8751 kcal/mol. The crystallographic water remains coordinated to Mg^{2+}. Additionally, the optimal nonbonded parameters between the attacking oxygen and donor sulfur were derived from Kollman (R^*_S = 2.0000 Å, ε_S = 0.2500 kcal/mol).

Charges were assigned using RESP, with two Na^+ counterions to neutralize charges, and the ligand was solvated with a 20 Å sphere of TIP3P water. The counterions were over 16 Å away from the active site and were fixed to remove potential artificial long-range electrostatic effects. Simulations were performed with a nonbonding cutoff of 16 Å, a time step of 1.5 fs with SHAKE, and at a constant temperature of 300 K using the Berendsen coupling scheme.

For the first minimization, the water molecules were relaxed for 1500 steps while all other atoms remained rigid. The water was then equilibrated using MD simulation for 20 ps. Next, we progressively removed the positional restraints on the protein-ligand complex (from 25 to 0 kcal/(mol $Å^2$)) over a period of 4000 minimization steps. The system then underwent MD equilibration for 20 ps. To avoid loss of water near the vicinity of the protein-ligand complex due to diffusion, the water shell was stripped and a new 20 Å shell of TIP3P was added, minimized, and equilibrated as described above. The protein-ligand complex was then minimized for another 4000 steps followed by an additional 20 ps of MD equilibration. A final minimization of 2000 steps was performed to yield our final structure.

2.3. QM determination of reaction barriers

We used three different models to study the methylation reaction between levodopa/dopamine and trimethylsulfonium (a substitute for S-adenosylmethionine). Model 1 was constructed by modeling only the ligand and trimethylsulfonium. Model 2 was constructed by taking Model 1 and adding four solvation layers of water, followed by minimization. Model 3 was generated from the minimized ligand pose generated using AMBER simulations of COMT with bound ligand (11,32). We included residues coordinated to Mg^{2+} (as well as its crystallographic water), the catalytic Lys144, and the regioselective Trp144 in our final construct for Model 3. In all models, the reaction coordinate is examined along the sulfur-oxygen axis.

Using a B3LYP functional, we performed a QST2 search for Models 1 and 3 to find the transition state of each methylation reaction (33). A mixed basis set of 6-31G for C, H and 6-311+G* for S, Mg, O, and N were used. Reactants of each methylation reaction were computed by geometry optimization. NBO analysis was used to examine the interaction energies between the two ligands. In addition, a level-of-theory study was performed on Model 1 to monitor the effects of changing density functionals on our barrier height calculations.

We calculate the energies of the transition state and reactant for Model 2 by taking the transition state for the simple model in the gas phase and solvating it with four water shells. The water was minimized using molecular mechanics, and then a single point calculation was done at the quantum level (using the same basis set) with the water treated using molecular mechanics. It is important to note that the single point calculations done in explicit solvent likely underestimate the stabilization due to the polarization of water since the dielectric response of

water is primarily determined by its orientational dynamics. Nonetheless Model 2 serves as an estimate of the methylation reaction occurring in an aqueous environment.

3. Results and discussion

3.1. Residues involved in regioselection are evolutionarily conserved

Our sequence analysis shows that residues comprising the hydrophobic wall of COMT are conserved throughout evolution (Figure 3.1). Many residues from the hydrophobic wall are as equally conserved as those coordinated to magnesium in the active site. Trp143 is as equally conserved as the catalytic base Lys144, though Trp143 also is known to initially bind to the adenosine moiety of SAM. For comparative purposes, we also show residues found in loops of the solvent accessible surface that expectedly have poorer sequence conservation.

High conservation of the hydrophobic wall residues suggests that its regioselective role is important for enzymatic function. Regioselection is most commonly associated with chemical specificity, but from a biological standpoint there is no need for such specificity since the products are degraded after interactions with COMT and monoamine oxidase (in no specified order) (1). The fundamental purpose of enzymes is to accelerate chemical reactions, and thus the conservation of regioselectivity has led us to hypothesize that COMT may have evolved regioselectivity as a method for chemical rate enhancement.

3.2. Reaction barrier differences between ligand conformations

The purpose of Models 1 and 2 is to deduce whether or not the ligands possess intrinsic regioselective properties, and what are the contributing components. Model 3 mimics the active site environment to give insight into what extent the protein contributes to regioselection. After obtaining the MD minimized structure for Model 3, we conducted DFT minimization using only the atoms described above. Our final complex after DFT minimization supports the hypothesis from Kuhn et al that Trp143 is involved in a cation-π with the catechol side-chain, as we can directly observe this interaction when modeling quantum mechanically (Figure 3.2).

From our calculations, we observe higher reaction barriers for the para conformation of levodopa and dopamine that effectively slows its rate of methylation (Table 3.2, Table 3.3). The energies calculated between Models 1 and 2 had no significant difference, likely due to utilizing a single-point calculation. Meta-methylation becomes more favorable for levodopa when occurring in the active site, potentially underlying an important role for Trp143 in minimizing the reaction barrier.

Our comparison of different functionals on Model 1 calculations unanimously shows that the meta-methylation reaction is more favorable (Table 3.3). Although there are large variations in how quantitatively accurate the barrier heights are, the qualitative trend clearly favors meta-methylation for both ligands. In our model of dopamine inside the active site of COMT (Model 3), there is actually a 4 kcal/mol preference for para-methylation which is in opposition to experimental observations. The meta/para ratio for dopamine is four times lower than the ratio for levodopa, and thus Model 3 of dopamine may be insufficient to detect such minor

differences. Nonetheless, from Models 1 and 2 it can be concluded that it is the enzyme itself that is giving rise to regioselection as the ligand itself has no preference.

3.3. Chemical differences between meta- and para- nucleophiles

The differences in reaction barriers alone do not give chemical insight into why regioselectivity takes place. To investigate further, we perform an NBO analysis to calculate the interaction energies between the methyl donor and ligand using Model 1. We find that the oxyanion in the meta TS is sp^3 hybridized while the para oxyanion is sp^2 hybridized. As a result, the interaction energy between the methyl group and the oxyanion is significantly higher in the meta conformation (Table 3.4).

Chemical descriptors can describe the consequences of differences in oxyanion hybridizations (34,35). The meta- nucleophile of each ligand is considered more electronegative and a harder base, indicative of it being a stronger nucleophile (Table 3.4). Notably, the meta-oxyanion of dopamine is still a more effective nucleophile than its para-oxyanion counterpart. Despite the increased nucleophilicity of the meta-oxyanion, the overall interaction energy between dopamine and trimethylsulfonium remains nearly identical for both conformations. Comparison of the total interaction energies between the ligand and methyl donor demonstrates that there is a significantly larger difference between the two conformations of levodopa as compared to dopamine (Table 3.4), consistent with experimental observations that levodopa has the higher meta-methylation preference (Table 3.1).

4. Conclusions

Our results propose that COMT utilizes regioselectivity not for the purposes of chemical specificity, but instead to enhance its catalytic rate by barrier minimization. Deprotonation of the 3' hydroxyl creates a more effective nucleophile and enables some ligands to confer regioselectivity as demonstrated by levodopa in our work. COMT has evolved to take advantage of this property by selecting for residues that dock ligands preferentially in the meta-conformation and thereby lowering the activation barrier for methylation.

Acknowledgements

This work was supported by the National Institute of Health Grant R01GM080742 and the ARRA supplement 3R01GM080742-03S1.

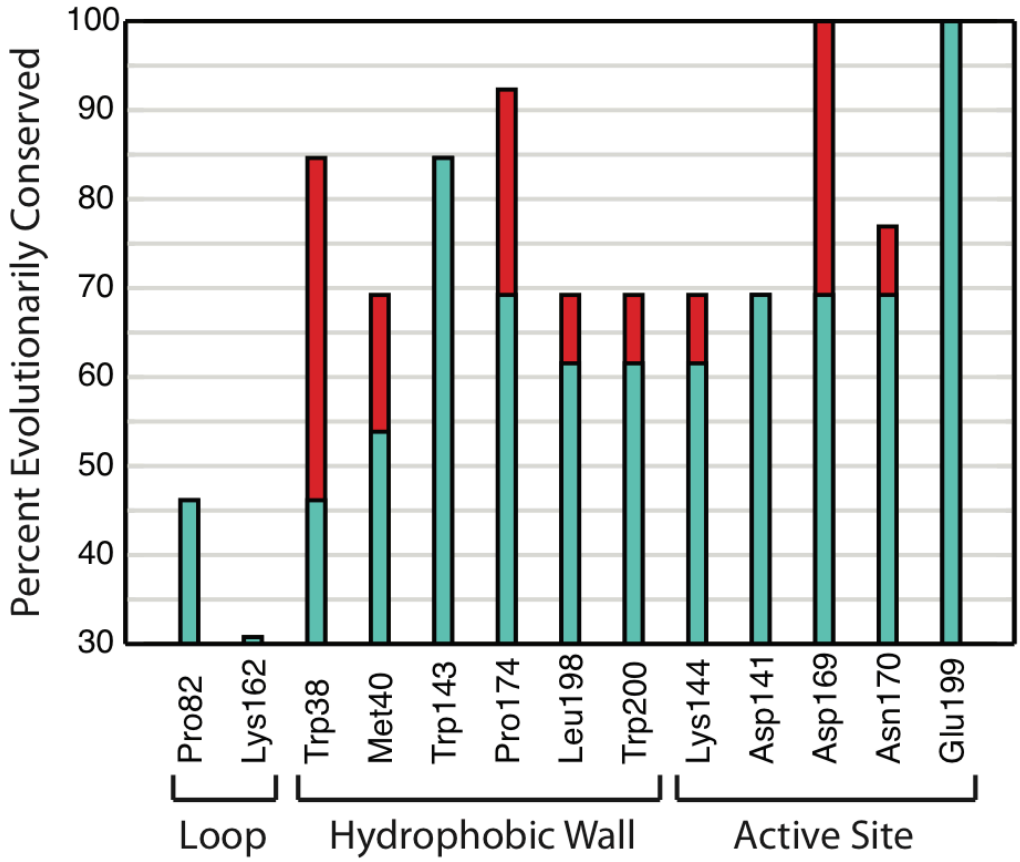


Figure 3.1. Sequence analysis of COMT shows high conservation of residues comprising hydrophobic wall. (Turquoise) Identical sequence conservation. (Red) Similar sequence conservation.

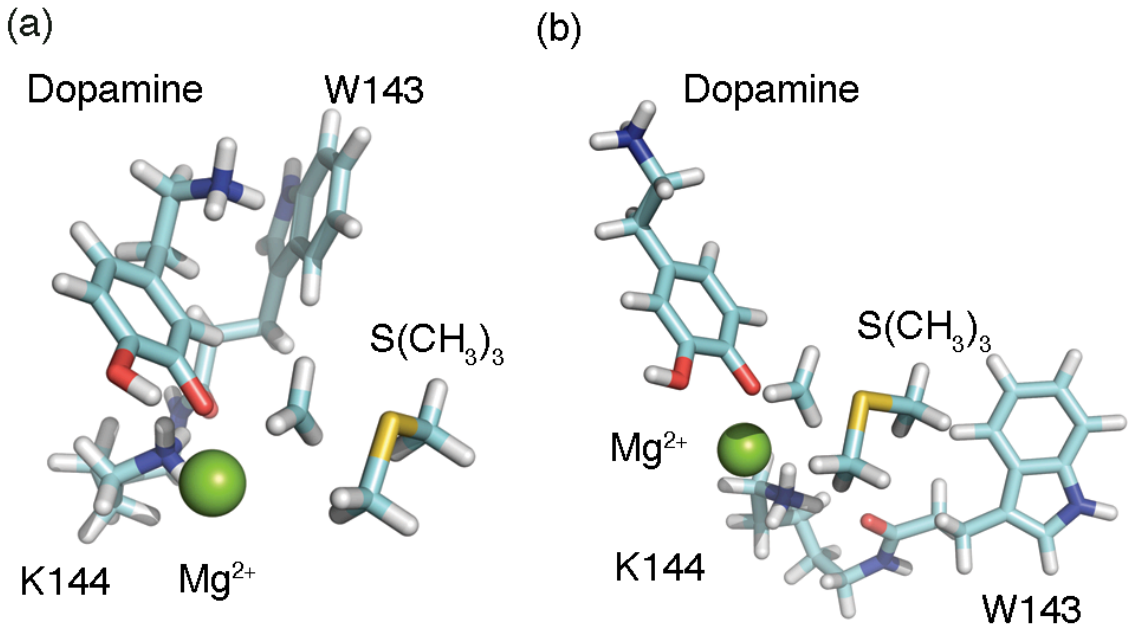
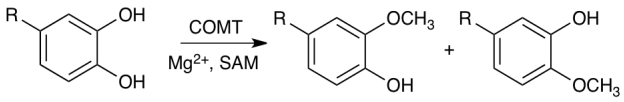


Figure 3.2. Transition state structures of COMT show Trp143 forming cation-π interaction exclusively in the meta conformation. Residues coordinated to Mg2+ removed for clarity. (a) Meta-methylation. (b) Para-methylation.

Tables

Table 3.1. Methylation of COMT Ligands



R	Ligand	Meta/Para Ratio (12)
$-CH_2CH_2NH_3^+$	Dopamine	4.2
$-CH_2CH(NH_3^+)COO^-$	Levodopa	15.4

Table 3.2. Reaction Barriers and Energy Decompositions for Different COMT Ligands Conformations[a]

Ligand-Conformation (Model)	ΔE_e	ΔE_{xc}	ΔT_S	ΔE
Levodopa-Meta (1)	11.9	33.2	-27.8	17.3
Levodopa-Para (1)	22.0	27.0	-28.9	20.1
Dopamine-Meta (1)	10.1	31.5	-24.8	16.8
Dopamine-Para (1)	30.4	26.6	-31.9	25.0
Levodopa-Meta (3)	37.9	36.8	-57.4	17.3
Levodopa-Para (3)	33.7	41.2	-51.9	23.0
Dopamine-Meta (3)	42.6	49.9	-58.3	34.2
Dopamine-Para (3)	39.5	42.8	-52.2	30.2

[a]Units are in kcal/mol. Legend: $\Delta E_e \equiv$ electrostatic energy, $\Delta E_{xc} \equiv$ exchange-correlation energy, $\Delta T_S \equiv$ kinetic energy, $\Delta E \equiv$ reaction barrier height.

Table 3.3. Performance comparison of difference density functionals on the transition state barrier height of the dopamine and levodopa reactions for the simplest model systems. Units in kcal/mol.

	Dopamine		Levodopa		Dopamine	Levodopa
	Meta-	Para-	Meta-	Para-	Meta-Para	Meta-Para
HF	58.57	66.19	59.41	62.14	-7.62	-2.74
BP86	34.38	42.97	34.96	39.24	-8.59	-4.29
BLYP	36.94	45.04	37.37	41.29	-8.09	-3.92
TPSSTPSS	36.71	45.25	37.02	41.82	-8.54	-4.80
MP2	40.67	47.57	41.06	45.65	-6.89	-4.60
SVWN5	27.55	36.36	28.01	33.05	-8.81	-5.05
VSXC	27.40	31.19	24.99	35.94	-3.79	-10.94
B3LYP	42.18	50.18	42.70	46.49	-8.00	-3.79
B3PW91	41.31	49.79	42.04	46.00	-8.48	-3.96
B972	42.79	50.97	43.57	47.37	-8.18	-3.79
O3LYP	43.91	52.11	44.90	48.14	-8.20	-3.24
BMK	43.39	51.02	44.02	47.57	-7.63	-3.55
BHandHLYP	48.49	56.29	49.11	52.45	-7.80	-3.34
M062X	43.46	50.65	43.66	48.43	-7.20	-4.76
tHCTHhyb	38.95	47.21	39.55	43.86	-8.26	-4.32
X3LYP	42.06	49.97	42.55	46.38	-7.91	-3.83
CAM-B3LYP	45.33	53.30	45.94	49.56	-7.98	-3.62
B2PLYP	41.21	57.95	41.59	45.73	-16.74	-4.14
mPW2PLYP	41.75	49.25	42.13	46.27	-7.51	-4.14
LC-wPBE	47.48	55.75	48.42	51.99	-8.27	-3.57
HSE2PBE	40.34	48.57	41.00	45.03	-8.23	-4.04
B1LYP	43.68	51.58	44.20	47.87	-7.90	-3.67
wB97X	44.53	52.04	45.25	48.88	-7.51	-3.62
wB97XD	42.03	49.68	42.64	47.13	-7.64	-4.49
LC-BLYP	48.13	56.20	48.89	52.30	-8.07	-3.41

Table 3.4. Conceptual DFT Descriptors and NBO Analysis for Model 1[a]

Ligand-Conformation	E_{int}	$E_{int}^{O\text{-}Me}$	χ	η
Levodopa-Meta	62.3	23.9	72.8	71.5
Levodopa-Para	37.9	2.7	64.0	55.8
Dopamine-Meta	48.9	26.7	70.9	70.9
Dopamine-Para	51.6	5.2	60.2	69.0

[a]Units in kcal/mol. Legend: $\chi \equiv$ electronegativity, $\eta \equiv$ chemical hardness, $E_{int} \equiv$ total interaction energy between SAM and ligand, $E_{int}^{O\text{-}Me} \equiv$ interaction energy between oxyanion and methyl group.

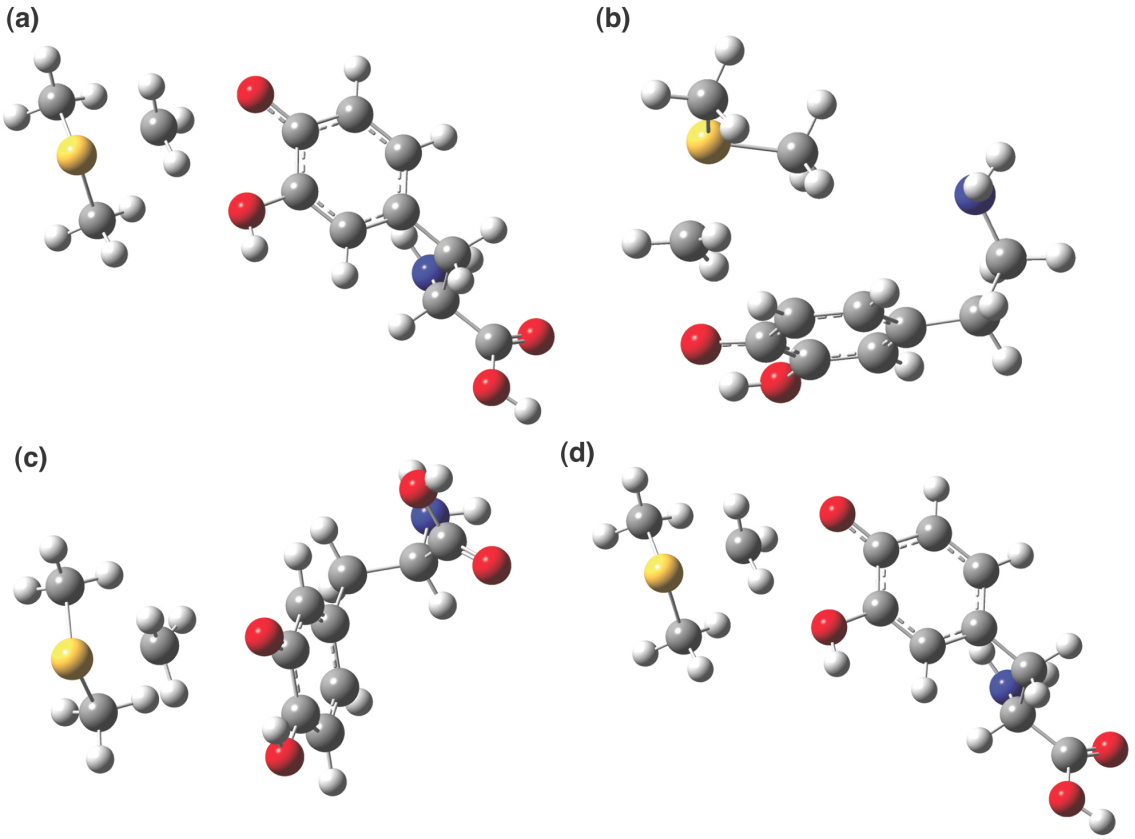


Figure 3.S1. Calculated transition states for Model 1 of Levodopa and Dopamine conformations. Model 2 uses the same transition states but in addition contains four additional solvation shells of water. Legend: hydrogen (white atoms), carbon (grey atoms), oxygen (blue atoms), and sulfur (yellow atom). (a) Dopamine Meta-Methylation. (b) Dopamine Para-Methylation. (c) Levodopa Meta-Methylation. (d) Levodopa Para-Methyation.

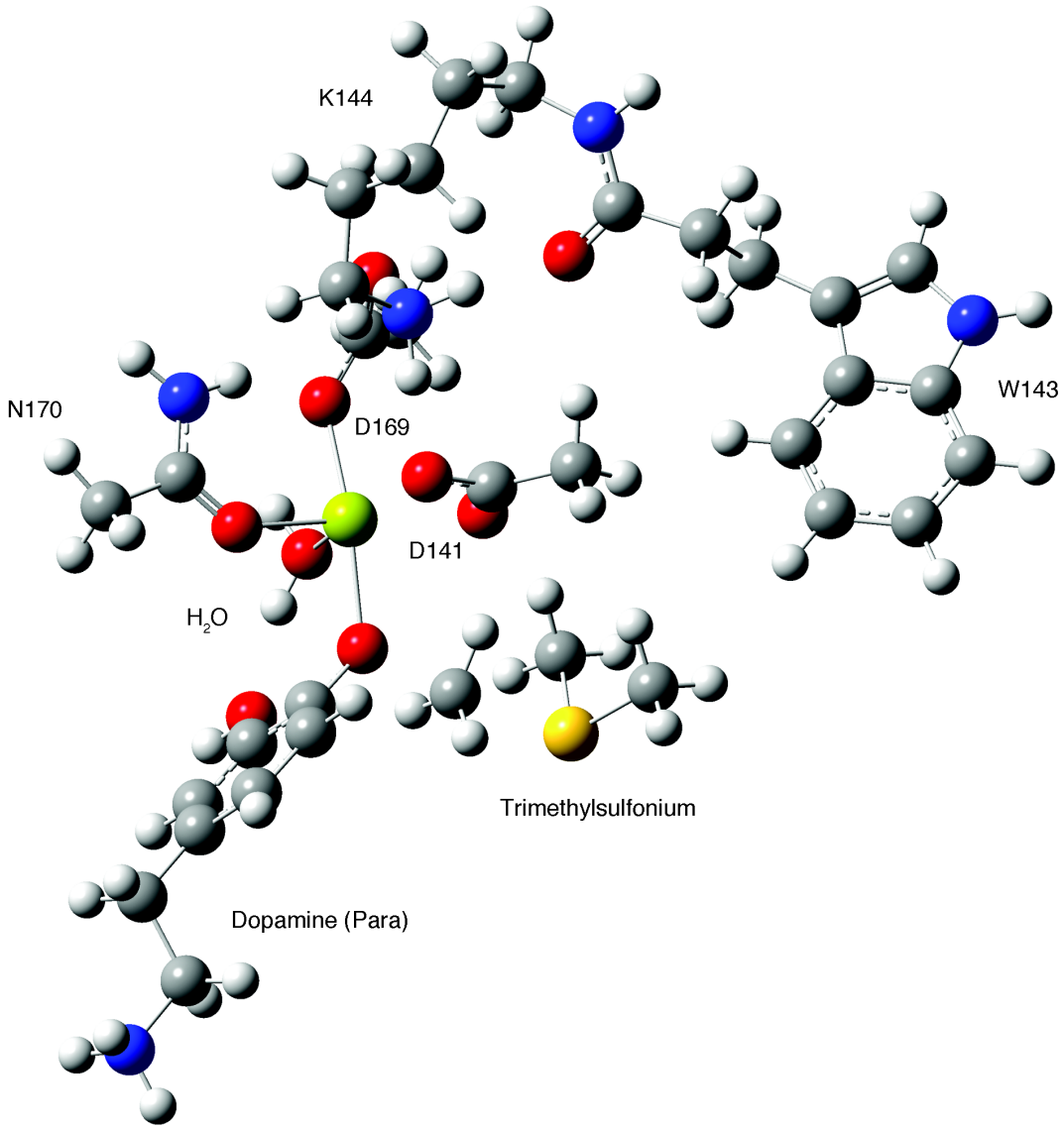


Figure 3.S2. Active site model of methylation. Dopamine methylated at the para position by trimethylfulonium is shown as an example. Legend: hydrogen (white atoms), carbon (grey atoms), oxygen (blue atoms), sulfur (yellow atom), and magnesium (green atom).

Conceptual DFT

Conceptual DFT is useful in obtaining qualitative chemical insight from quantum chemical calculations (34,35). In this work we have considered several chemical descriptors. Chemical hardness is traditionally defined as the difference between an atom's ionization potential and electron affinity. Applying Koopman's Theorem, we can approximate the hardness in DFT as

$$\eta = \varepsilon_{LUMO} - \varepsilon_{HOMO}$$

which is simply the energy difference of the HOMO-LUMO gap. Similarly, we can then define the electronegativity as

$$\chi = -\frac{1}{2}\left(\varepsilon_{HOMO} + \varepsilon_{LUMO}\right)$$

Traditionally the energy in DFT has been decomposed as

$$\Delta E = \Delta T_S + \Delta E_e + \Delta E_{xc}$$

where TS, Ee, and Exc are the noninteracting kinetic, electrostatic, and the exchange-correlation energies respectively. Alternatively, we can decompose the energy difference as

$$\Delta E = \Delta E_s + \Delta E_e + \Delta E_q$$

where the individual differences arise from steric, electrostatic, and quantum contributions respectively. The steric contribution arises from the Weizsäcker kinetic energy

$$E_s \equiv T_W = \frac{1}{8}\int \frac{\left|\nabla\rho(\mathbf{r})^2\right|}{\rho(\mathbf{r})} d\mathbf{r}$$

and the quantum contribution is

$$E_q = E_{xc} + T_S - T_W$$

CHAPTER IV: DISRUPTIVE MRNA FOLDING INCREASES TRANSLATIONAL EFFICIENCY OF CATECHOL-*O*-METHYLTRANSFERASE VARIANT

Douglas Tsao[1,2], Svetlana A. Shabalina[3], Josée Gauthier[2], Nikolay V. Dokholyan[2,4,*], and Luda Diatchenko[2,*]

[1]Department of Chemistry

University of North Carolina, Chapel Hill, NC 27599, USA;

[2]Center for Neurosensory Disorders, School of Dentistry

University of North Carolina, Chapel Hill, NC 27599, USA;

[3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA;

[4]Department of Biochemistry and Biophysics, School of Medicine

University of North Carolina, Chapel Hill, NC 27599, USA

Corresponding Authors: Luda Diatchenko (lbdiatch@email.unc.edu) and Nikolay V. Dokholyan (dokh@med.unc.edu)

ABSTRACT

Catechol-*O*-methyltransferase (COMT) is a major enzyme controlling catecholamine levels that plays a central role in cognition, affective mood and pain perception. There are three common *COMT* haplotypes in the human population reported to have functional effects,

divergent in two synonymous and one nonsynonymous position. We demonstrate that one of the haplotypes, carrying the non-synonymous variation known to code for a less stable protein, exhibits increased protein expression *in vitro*. This increased protein expression, which would compensate for lower protein stability, is solely produced by a synonymous variation ($C^{166}T$) situated within the haplotype and located in the 5' region of the RNA transcript. Based on mRNA secondary structure predictions, we suggest that structural destabilization near the start codon caused by the T allele could be related to the observed increase in COMT expression. Our folding simulations of the tertiary mRNA structures demonstrate that destabilization by the T allele lowers the folding transition barrier thus decreasing the probability of occupying its native state. These data suggest a novel structural mechanism whereby functional synonymous variations near the translation initiation codon affect the translation efficiency via entropy-driven changes in mRNA dynamics and present another example of stable compensatory genetic variations in the human population.

INTRODUCTION

Catechol-*O*-methyltransferase (COMT) deactivates neurotransmitters and metabolizes catechol-containing structures by methylation of a hydroxyl group (12). The implications of COMT activity are broad and can influence factors such as general cognitive function (24,36,37), addiction (25), stress response (25), and pain sensitivity (13). Three genetic variants of COMT have been identified in the human population corresponding to low, average, and high pain sensitivity haplotypes (LPS, APS, and HPS) (13). Higher COMT activity corresponds to lower pain sensitivity and vice versa. A silent mutation differentiates between low (LPS) and high

(HPS) pain sensitive phenotypes via reduced HPS protein levels (15), while APS is characterized by a valine to methionine substitution at amino acid position 108 that reduces its intrinsic activity through lowering protein stability (12,16) (Fig. 4.1A). These haplotypes have also been associated with risk of fibromyalgia (18), temporomandibular joint disorder (TMJD) (13), postsurgical pain (38,39), responses to drugs (20) and development of brain white matter (21).

The ability of highly structured regions of mRNA to inhibit protein expression was recognized for a long time (40-42). However, the exact mechanisms of this inhibition and its relative contributions to regulation of translation efficiency in live cells have only limited examples (43,44). Thus, several *in vitro* studies have shown that RNA transcripts containing extremely stable stems with melting temperatures higher than 70°C can decrease protein expression at the level of ribosomal translocation (45). The underlying factor preventing translation at highly stable regions is thought to be the ribosome itself. It has been shown that the ribosome contains an intrinsic helicase activity allowing it to read the individual bases (45). Thus, RNA motifs that are too difficult to unwind cause the ribosome to stall on the transcript.

Protein synthesis is highly regulated at the initiation stage enabling rapid, reversible and spatial control of gene expression (46-49). Prokaryotic translation of mRNA is regulated at both the 5' and 3' ends of a transcript during initiation(50). For eukaryotes, initiation of translation proceeds by the ribosome scanning from the 5' end of the transcript to the initial start codon (41,51). Scanning through the transcript is facilitated by the eIF4 factor unwinding structured RNA regions through an ATP-dependent process (40), and because of the scanning mechanism ribosomes cannot bind circular mRNA transcripts (52). Earlier work has demonstrated that gene expression can be repressed by increasing the stability of 5' end mRNA secondary structures (53). Recent experiments with GFP constructs have also shown that the folding free energy of

the 5' end of an mRNA transcript is most correlated with protein expression, as opposed to a codon bias (54). Furthermore, reduced stability of the mRNA at the translation-initiation site was found to be a common feature for most species (55).

To uncover the translation mechanisms that allelic variants of common COMT haplotypes contribute to variation in COMT activity, we performed a set of molecular and computational studies. We first conducted *in vitro* translation studies of three haplotypes in rabbit reticulocyte lysates. Unlike the *in vivo* expression system, we did not observe a difference in an amount of translated COMT protein between LPS and HPS haplotypes, suggesting that rs4818-dependent stem-loop structure (15) requires additional cellular chaperons to affect translation efficiency. However, we observed robust increase in amount of protein of APS haplotype-coded mRNA. Here we show how APS haplotype-specific T allele of the SNP rs4633 located at the 5' end of mRNA near the ribosomal binding site, rather than non-synonymous *met*^{158} variation, modulates protein expression *in vitro*. We also conduct secondary structural analysis and perform simulations at the 5' end of each haplotype using discrete molecular dynamics to determine the mechanism by which the T allele at rs4633 alters translational efficiency (46,56,57). Our results reveal a novel mechanism by which the dynamics of mRNA structures near the initial start codon may influence efficiency of translation initiation.

MATERIALS AND METHODS

In Vitro Translation

COMT cDNA coding for three haplotypes and LPS-T^{166} mutant were cloned into a pCMV-Sport6 vector as described previously (15). The mRNA templates used for translation were generated by first restriction enzyme digestion using HindIII to create a linear plasmid.

Digested plasmids were subsequently cleaned up using a PCR purification kit (Qiagen). *In vitro* transcription was performed by adding SP6 RNA polymerase (Promega) along with rNTPs and incubated in a reaction buffer under conditions provided by the manufacturer. RNA was purified from the mixture using Trizol (Invitrogen) and subsequently dissolved in water. The RNA integrity was evaluated by running the samples on the Bioanalyzer 2100 (Agilent).

The *in vitro* translation reaction was carried out using 1 μg RNA template, 17.5 μL rabbit reticulolysate, 0.5 μL amino acid mixture (-Met), 1 μL ^{35}S-labeled Methionine (1200 Ci/mmol), 0.5 μL RNasin, and diluted to a total reaction volume of 25 μL. To denature the RNA we heat up the samples for 3 minutes at 70°C and immediately place on ice. For RNA secondary structure formation, we heat denature then subsequently add 5 mM $MgCl_2$ and cool at a rate 0.1°C/sec to a final temperature of 15°C. Once the RNA template is added to the rabbit lysate mix, we incubate for 1.5 hours at 30°C. The reaction is stopped by adding 1x Laemmli buffer and heating for 4 minutes at 80°C.

We quantified the amount of protein product by separating via SDS-PAGE. The gel is initially placed in fixing solution (50% methanol, 40% water, 10% acetic acid) for 30 minutes under gentle rotation. Afterwards the gel is soaked in a rinsing solution (85% water, 7% methanol, 7% acetic acid, 1% glycerol) for 5 minutes with gentle rotation. The gel is then placed in a drier with vacuum pump for 1.5 hours at 80°C. The gel is then placed in a cassette with PhosphorImager screen and later quantified using Storm PhosphorImaging System (Molecular Dynamics).

To verify that our radiolabeled protein product is COMT, we performed immunoprecipitations on several lysate reactions. After *in vitro* translation reaction, an equal amount of NET buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 7.4) is added. We

use Ultralink Protein A/G agarose beads and equilibrated them by washing with 0.5 mL NET buffer twice per 100 μL beads and resuspending in 100 μL NET buffer. For each lysate reaction, 5 μL of primary COMT antibody were added and incubated overnight with rotating at 4°C. Then 50 μL of equilibrated Protein A/G agarose beads are added and incubated at 4°C for 4 hours. Samples are then centrifuged for 5 minutes to remove the supernatant. The supernatant is saved for further SDS-PAGE analysis. The beads are then subsequently washed with 50 μL NET buffer twice by rotating for 5 minutes in 4°C. The proteins are removed from the beads by dissolving them with 25 μL of Novex Tris-Glycine SDS solution and boiled for 4 minutes at 80°C. The supernatant from the boiling reaction contains our immunoprecipitated protein and is analyzed via SDS-PAGE.

Transfection and Western Blotting

The cDNA clones coding for three COMT haplotypes were transfected into mammalian cell lines as described previously (15). COS-1, Hek-293, HepG2, and MCF-7 cell lines were purchased from ATCC and maintained in media (DMEM with 10% FBS, 4.5g/L glucose, L-glutamine, and sodium pyruvate for COS-1, Hek-293, and HepG2; RPMI 1640 with 5% FBS and L-glutamine for MCF-7) in accordance with manufacturer's recommendations. The Western blotting was performed as described previously (15) using anti-hCOMT antibody derived from rabbit (Chemicon, ab5873).

Secondary Structure Analysis of COMT Allelic Variants

COMT allelic variants and randomly generated sequences were computationally 'folded' and the predicted minimum free energy of the secondary structure was calculated for different

window sizes, using our implementation of the algorithm described by Zuker (57). Energy minimization was performed by dynamic programming method using an improved algorithm for evaluation of internal loops (58).

We estimated the free-energy penalty associated with breaking (opening) of the target's local secondary structure (target structure opening, ΔG kcal/mol), considering local disruption of secondary structure in windows with different lengths. Free-energy changes were approximated with nearest-neighbor free energy parameters using the program OligoWalk (59). Here, we consider local structure for a set of suboptimal structures (Fig. 2A). Each structure contributes to the free energy penalty for disruption of structure in proportion to a Boltzmann weight; and the summations over all suboptimal structures were provided. Thus, this difference between the free energy of each suboptimal structure and the free energy of the corresponding suboptimal structure without base pairs in the region of complementarity in 30-nt window length is defined as the energy required for target structure opening (Fig. 2A). Monte Carlo simulation and analysis of randomized sequences (47,60) was used for estimation of the significant difference between target structure opening, ΔG, of two COMT allelic variants. One-thousand unique random sequences for each allelic variant were generated by shuffling the first 210 nucleotides of the COMT mRNA sequence and iteratively mutating two positions randomly along the sequence (except position 166). Each generated sequence is then checked to verify that the %GC and %AU content remain identical to the original COMT gene. Once the sequences are cross-checked with one another to ensure there are no duplicates, we create two sets of sequences where position 166 is occupied by either a C or U. The free-energy penalty associated with opening of the target's local secondary structure (ΔG kcal/mol) for all random sequences, considering local disruption of secondary structure in windows with 30-nt length was calculated.

P-values for randomizations and for difference between the C and T alleles were determined by paired *t*-tests.

Discrete Molecular Dynamics

Traditional molecular dynamics simulate the motions of particles by solving Newton's equations of motion for a defined system using an integration algorithm. In DMD, simulations proceed according to the conservation laws of energy, momentum, and angular momentum and are evaluated as a series of two-body interactions. The efficiency of the engine is based on an algorithm that searches through an event table, where velocities are only modified as necessary. Here we classify an event as the instance in which two particles are within a defined interaction range as defined by their potential. The potentials used in DMD are discretized to accommodate the discontinuous nature of the simulations.

Let us consider a system of particles, including two particles i and j both of mass m that occupy initial positions of r_{i0} and r_{j0} with initial velocities of v_i and v_j. For simplicity we discuss a scenario where the interaction potential consists of one step, otherwise known as a square well potential, defined as

$$U_{ij} = \begin{cases} \infty, & r < \sigma_1 \\ -\varepsilon, & \sigma_1 < r < \sigma_2 \\ 0, & r > \sigma_2 \end{cases}$$

The variable σ_α defines an interaction distance, where $\alpha = 1$ refers to the hard sphere repulsion and $\alpha = 2$ is the attractive interaction. During a simulation, a table is generated by calculating event times for each pair of particles. The time interval in which an event may occur between two particles is

$$t_{ij}^{(\alpha)} = \frac{-b_{ij} \pm \left[b_{ij}^2 - v_{ij}^2\left(r_{ij}^2 - \sigma_\alpha^2\right)\right]^{1/2}}{v_{ij}^2}$$

where

$$\mathbf{r}_{ij} = \mathbf{r}_{i0} - \mathbf{r}_{j0}$$

$$\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$$

$$b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{v}_{ij}$$

and the plus-minus sign refers to two particles either approaching or receding from each other, respectively. The trajectory of particle *i* is evaluated with respect to time as

$$\mathbf{r}_i\left(t + t_u\right) = \mathbf{r}_i(t) + \mathbf{v}_i(t)t_u$$

where the time unit t_u is determined by comparing the shortest event time to the maximum allowed time interval t_m. If $t_{ij}^{(\alpha)}$ is greater than t_m, then the particles are only permitted to move for t_m and a new table is subsequently generated. Otherwise the shortest event time is the first considered, and the velocity changes according to the conservation laws of energy and momentum.

When particles *i* and *j* are determined to interact at $t_{ij}^{(\alpha)}$, the squared difference in the particles' positions is compared to the squared interaction distance prior to a change in the potential (i.e., before $r_{ij}^2 - \sigma_\alpha^2 = 0$). There are three possible scenarios for our one-step potential as the two particles approach one another. During an attractive encounter, if $r_{ij}^2 > \sigma_2^2$ then the magnitude of separation between the particles decreases and the change in velocities will be

$$\Delta\mathbf{v}_i = -\Delta\mathbf{v}_j = \frac{-\mathbf{r}_{ij}}{2\sigma_2^2}\left[b_{ij} + \left(b_{ij}^2 - \frac{4\sigma_2^2\varepsilon}{m}\right)^{1/2}\right]$$

However if $r_{ij}^2 < \sigma_2^2$, then the two particles increase their magnitude of separation and the change in velocities will be

$$\Delta \mathbf{v}_i = -\Delta \mathbf{v}_j = \frac{-\mathbf{r}_{ij}}{2\sigma_2^2}\left[b_{ij} - \left(b_{ij}^2 + \frac{4\sigma_2^2\varepsilon}{m}\right)^{1/2}\right]$$

In the case of hard sphere repulsions, the change in velocities is simply

$$\Delta \mathbf{v}_i = -\Delta \mathbf{v}_j = \frac{-\mathbf{r}_{ij} b_{ij}}{2\sigma_1^2}$$

After each event, we remove from the table events that were calculated with the previous two particles since their velocities and positions have changed. The number of possible events with these two particles are then recalculated and sorted within the table. Simulations then proceed through the table, where particles are allowed to move between time intervals, until either $t_{ij}^{(\alpha)} < t_m$ or the table of events is depleted.

We incorporate the Andersen thermostat to simulate under canonical (constant *N*, *V*, *T*) conditions (61). Temperature is maintained constant by surrounding the system of particles with a heat bath. The heat bath itself is comprised of imaginary ghost particles (62,63) with number density ρ_g that undergo stochastic collisions with the system via a Poisson process

$$P(t) = q\exp[-qt]$$

Here $P(t)$ represents the probability that a randomly chosen particle within the system undergoes a collision with a ghost particle at time t. The constant q represents the rate at which system particles undergo collision with the ghost particles. This may also be referred to as the heat exchange rate, which is determined by

$$q = \rho_g \sigma_1^2 \left(\frac{16\pi kT}{m}\right)^{1/2}$$

The momentum of a particle after collision with a ghost particle at t_u is selected randomly from a Boltzmann distribution of values at temperature T.

Three-bead RNA Model

We perform the RNA folding simulations using a simplified 3-bead model (freely available on the web at http://ifoldrna.dokhlab.org/) (56). For each nucleotide, each bead represents a phosphate, sugar, and base. Interactions contained in the model include standard Watson-Crick base pairing, G-U base pairing, base stacking, phosphate-phosphate repulsion, hydrophobic interactions, and loop entropy.

Replica-Exchange Simulations and Analysis

To ensure adequate sampling of the conformational space of each haplotype, we utilize replica-exchange where R replicas are simulated each at a temperature T_i where i represents the index of that particular replica (64). A random walk in temperature space is performed by exchanging the temperatures between two replicas i and j under the probability

$$p = \begin{cases} 1, & \text{for } \lambda \le 0 \\ \exp(-\lambda), & \text{for } \lambda > 0 \end{cases}$$

where

$$\lambda = \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j} \right) \left(E_j - E_i \right)$$

and E_i is the total energy of the system within the ith replica. Thus as the macromolecule explores conformational space, the energy changes accordingly. Swapping the temperatures between replicas allows conformations that are stuck in local energy minima to escape and resume exploring other conformations. For our RNA folding simulations, we performed replica-

exchange simulations with nine replicas (T = 0.1, 0.15, 0.2, 0.225, 0.25, 0.275, 0.35, 0.4, and 0.5 ε/k_B) per allelic variant for 2 x 10^6 time steps. Energies of each RNA conformation throughout the simulation are evaluated according to parameters published previously (56).

Simulations can be analyzed using the weighted histogram analysis method (WHAM) to determine various thermodynamic quantities by deriving a partition function using the trajectories (65). We compute the specific heat of folding using

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B^2 T^2}$$

where < E > is the average potential energy and < E^2 > is the average squared potential energy of the system for each specific temperature (62).

Contact maps provide a useful measure of gauging the frequency of certain conformations over the simulation. For our purposes, we are interested in the frequency of base pair formations and thus have limited our contacts to the base beads (as opposed to including the sugar and phosphate beads). We define a base pair between two atoms i and $j > i+3$ as less than 6.5 Å. From the contact maps, we can compare structures derived from simulations to secondary structure prediction programs by comparing contact maps of the former to dot plots of the latter. We can also deduce the secondary structures from the tertiary structures by calculating base pairs formed for each trajectory according to the parameters in the force field and determine which secondary structures are most probable (56).

We have clustered conformations using the OC hierarchical clustering package (available at http://www.compbio.dundee.ac.uk/downloads/oc). RNA structures derived from our simulation trajectories were clustered according to RMSD. The lower the RMSD between two RNA structures, the less distant (in terms of clustering) they are from one another. The OC algorithm works by first taking two structures that have the minimum distance and assigning

them as a cluster, and then search across all other structures and comparing their distances amongst other structures. Structures that have been clustered together are considered a single entity. Representative tertiary structures were derived from three clustering methods: single (where the minimum distance between the two clusters is taken as the distance), complete (where the maximum distance between the clusters is taken as the distance), and means (where the average distance between clusters is taken as the distance). Structures that were found to be most dominant for all three clustering methodologies were considered most representative.

RESULTS

In vitro translation of COMT haplotypes

To study the precise molecular mechanism(s) whereby the mRNA of COMT haplotypes (Fig. 4.1) produce different protein levels, we first employed an *in vitro* translation approach that is very effective in isolating putative mechanisms involving differential effects at the ribosomal level. We performed an endpoint kinetics assay using rabbit reticulocyte lysates. The advantage of this *in vitro* system is that external biological factors that regulate protein synthesis are absent. We found that the APS haplotype demonstrates higher protein expression levels compared to both the LPS and HPS haplotypes while LPS and HPS haplotypes have equivalent expression levels *in vitro* (Fig. 4.1B).

There are two unique alleles an APS haplotype carries within its transcribed region – a T allele of the SNP rs4633 at the 5' end of the mRNA (+32 nt downstream from the start codon) and an A allele of the SNP rs4680 within the second exon of the gene (Fig. 4.1A). As we previously showed, the structural mRNA differences within the second exon are the most

pronounced between LPS and HPS haplotypes (15), not the APS haplotype. Consequently, we concluded that it is unlikely that SNP rs4680 contributes significantly to the protein levels in the *in vitro* translation experiment. In contrast, SNP rs4633 is situated at the 5'end of the COMT mRNA near the start codon; a region which showed the strongest association between stability of mRNA folding and the rates of translation initiation expression levels of individual genes (54,66). To test the individual contribution of SNP rs4633 in the increase in protein levels observed for APS haplotypes, we created a C to T mutant at position 166 for LPS haplotype (LPS-T^{166}). Our results show that mutating LPS at position 166 from C to T recapitulates high expression levels characteristic for APS (Fig. 4.1B). Thus the determining factor for translation efficiency of COMT resides in the SNP rs4633 alone.

Secondary structure prediction of 5' region

We then studied local RNA secondary structures contributing to effect of SNP rs4633. It was shown that free energy stability of the 5' region of an mRNA transcript is correlated with translational efficiency (47,48,54,66-68). Transcripts that have less stable RNA structural elements near the 5' end have higher translation rates (54), presumably because tight binding to the initial start codon becomes difficult for the ribosome initiation machinery (46,49,51,54). To test if the T allele of rs4633 specific for APS haplotype results in a change in free energy, we initially utilize secondary structure prediction programs to calculate the free energy of the 5' end for different respective RNA.

We predicted mRNA local secondary structures in the vicinity of the SNP rs4633 and the start codon by employing different algorithms for both variants (56,58,69). Comparison of the

optimal structures for both C and T-allelic variants shows that the main structural differences due to the SNP rs4633 lies within the structural regions of Loop I, Stem I and Stem II (Fig. 4.2A; Supplementary Material, Figs. 4.S1-4.S3). Loop I of the ^{166}T variant (nt 119-126) is more flexible with an additional two nucleotides (nucleotides 126 and 127) comprising this single-stranded region. The two nucleotides present in Loop I of the ^{166}T variant causes Stem I to lose two Watson-Crick base pairs and lower the stem's stability. Consideration of these two regions alone, the T allele-carrying mRNA is less stable. However, both free energy calculations predict that its structure at Stem II downstream from the start codon is ~1 kcal/mol more stable than the wild type. This is primarily due to an additional A-U base pair formed unique to the T allele-carrying mRNA in Stem II between nucleotides 156 and 165 that increases its stability from an enthalpic standpoint. From visual inspection, it can also be seen that the additional unpaired bases found in the hairpin loop of Stem II for ^{166}C variant are entropically disfavorable due to the number of single-stranded nucleotides comprising the terminal loop compared to only four base pairs within the stem.

To verify whether this structural region is stable independent of the surrounding sequence, we truncated the sequence near the neighboring junctions (nucleotides 119 through 171) and refolded the structures using Mfold, Afold and RNAstructure (57,58,69). All programs predict the same optimal local structures as the 210 nt length transcripts (Fig. 4.2A). Furthermore, we predicted all suboptimal structures for both C and T allelic variants when percent suboptimality is set to 30 (when only folding within 30% from the minimum free energy will be computed). Identical Stem I loops surrounding AUG codons were found in two most stable RNA local structures with the total free energies (ΔG) of -14.8 and 8.63 kcal/mol in the ^{166}C-allelic variant, respectively (Fig. 4.S4). The ^{166}T-allelic variant produced four structures

with the energies ranging from ΔG=-16.6 kcal/mol to ΔG=-10.05 kcal/mol, only one of those identical at Stem I loop (Fig. 4.S5). Thus, there are not only differences in the secondary structures between ^{166}C and ^{166}T allelic variants, but ^{166}T-allelic variant also produces a higher diversity of suboptimal structures.

We then estimated the level of pairing and free energy of target breaking (opening) for the ^{166}C-allelic variant mRNA and the ^{166}T variant using full length transcripts and truncated transcript sequences of different lengths starting from 210 nucleotides, where approximately a half of the sequence length is located in the 5'UTR and the second half is in the coding region of the COMT gene. We modeled the dynamic process of transcript folding and target breaking using 30 nucleotide windows (59,70) (see **Materials and Methods**). Profiles of the free energy of target breaking for ^{166}C and ^{166}T variants (Fig. 4.2B) show that mRNA secondary structures in the vicinity of the start codon (30-nt length window) are less stable and the free energy of target breaking is significantly ($P = 0.0016$) higher for the ^{166}T variant of rs4633 (specific for APS haplotype) relative to the ^{166}C variant of rs4633 (specific for LPS and HPS haplotypes). On the other hand, Monte Carlo simulation of the sequences in the vicinity of the start codon showed that the differences in free energy of mRNA secondary structure target opening for the ^{166}T variant and the ^{166}C variant of rs4633 were not random ($P < 0.05$). Consistent with these findings, our secondary structural analysis using Mfold, Afold and RNAstructure (57,58,69) also demonstrates that the allelic variation of rs4633 directly affects the RNA structure surrounding the start codon (Fig. 4.2B). Thus, our secondary structure predictions thus provide an important insight that there is an independent motif in which there are structural differences for each haplotype.

Simulations predict different folding barriers

Our folding prediction and analysis of mRNA local secondary structures revealed that the ^{166}T allele promotes base pair disruption near the vicinity of the start codon (Fig. 4.2B). The lower free energy stability of the ^{166}T-allelic variant implies that there exists a lower energy barrier separating the folded and unfolded states. The eukaryotic initiation factor eIF4a facilitates translation in an ATP-dependent manner by unwinding RNA secondary structure to enable ribosomal translocation (40). Thus if there are differences in the energetic barriers between the folded and unfolded states for the haplotypes, then structures with a lower energy barrier height would undergo more efficient translation since there would be a higher probability for the structure to exist in an unfolded conformation (71,72). To test this hypothesis, we generate tertiary structures of the RNA motifs by simulating the dynamics **(**Fig. 4.3A,B**)** of each allelic variant using discrete molecular dynamics (DMD) (56).

Simulations were performed for both ^{166}C and ^{166}T-allelic variations of rs4633 of COMT transcripts at the 5' region between nucleotides 119 through 171 using an RNA three-bead model (56). Since we observe multiple transitions between the folded and unfolded states, there is adequate sampling to enable determination of the thermodynamics of the folding transition using the weighted histogram analysis method. A comparison between the ^{166}C and ^{166}T alleles of mRNA transcripts shows that the peak denoting the folding transition temperature is slightly higher for the ^{166}C allele, demonstrating its increase in thermodynamic stability (Fig. 4.3C).

We find that both alleles adopt a native conformation (Fig. 4.3A,B) that is in line with secondary structural predictions (Fig. 4.2A). Both the C and T alleles fold into their respective native conformations at -25.1 kcal/mol and lower (Fig. 4.3D). Notably, the ^{166}C allele has a higher probability in existing in this low energy state. The most stable structures that are unique

to the ^{166}C allele are formed due to transient base pair formations in the loops of Stem I and Stem II (Fig. 4.3D). In contrast, the ^{166}T allele-carrying mRNA is more likely to adopt conformations at higher energies along its folding pathway (Fig. 4.3D); a consequence of its lower folding transition temperature (Fig. 4.3C).

Since we know that the T allele is responsible for disrupting the local secondary structure near the start codon (Fig. 4.2B), we wanted to deduce the destabilizing effects on the overall tertiary structure of that region. From our simulations, we determined the flexibility of each allelic variant's tertiary structure by calculating the root mean square fluctuation (RMSF) of the native ensemble (as highlighted in Fig. 4.3D). We find that the dynamics of ^{166}T allele-carrying mRNA are highly entropy driven (Fig. 4.3B; right most structure) with RMSFs up to 11 Å for a single nucleotide (Table 4.S1). In contrast, the ^{166}C allele-carrying mRNA has a RMSF less than 6.5 Å for all nucleotides (Fig. 4.3A; Table 4.S1). The contact map highlights the attempts made by ^{166}T allele-carrying mRNA to fold into its native structure (Fig. 4.3E,F). The plethora of contacts demonstrates the competing states that lead to a rugged energy landscape **(**Fig. 4.3D). Thus, examining the dynamics of ^{166}T allele-carrying mRNA suggests that the free energy barrier height between folded and unfolded states is smaller (Fig. 4.3C) and therefore likely to adopt higher energy states. Consequently, the ^{166}T allele exists in an unfolded intermediate state more frequently than the ^{166}C allele-carrying mRNA (Fig. 4.3D) (71-73).

Haplotype-dependent COMT Protein Expression in Mammalian Cell Lines

To examine if the higher efficiency of protein translation contributed by the ^{166}T allele plays a substantial role at the cellular level, we carried out a series of transfection studies. As reported previously, in PC12 rat pheochromacytoma cells, COMT protein expression levels is

reduced 25-fold in the HPS haplotype; however, LPS and APS haplotypes display comparable protein levels (15). To determine whether this effect is a general feature of COMT protein expression in mammalian cells or specific to the PC12 cell line, we transfected expression vectors with the three COMT haplotypes in a number of different cell line with divergent tissue origin: COS-1 monkey kidney cells, HEK-293 human embryonic kidney cells, HepG2 human liver cells, and MCF-7 human breast cancer cells. We find that all transfected cell types consistently exhibited the same qualitative trend in protein expression (Fig. 4.4), where LPS exhibits higher protein expression than HPS, and HPS shows the most reduced protein expression. Notably, APS showed comparable protein levels to LPS in COS-1 and HepG2 cell lines but not in HEK-293 and MCF-7 where APS showed the highest protein level **(**Fig. 4.4). Thus, it appears that an increase in translation efficiency of the APS haplotype does contribute to cellular protein levels and this contribution is tissue-specific.

DISCUSSION

^{166}T mutation contributes to higher translation rates

Our *in vitro* translation data demonstrate that the APS haplotype of COMT has a higher rate of translation compared to LPS and HPS, while LPS and HPS haplotypes show similar level of expression (Fig. 4.1B). Furthermore, our results suggested that the upstream 5'-end structure in allelic dependent manner largely controls the *in vitro* translation rate. The APS-specific ^{166}T allele of rs4633 was hypothesized to drive the difference in protein translation efficiency. The ^{166}T allele of rs4633 was also a strong candidate for the translational "switch" because of its

unique location near the start codon, the RNA area known to contribute the most to the RNA structure-dependent translation initiation rate (54). To test this hypothesis, we created a LPS mutant carrying the ^{166}T allele of rs4633 (LPS-T^{166}) specific for APS haplotype. This mutant has similar protein expression levels in comparison to APS (Fig. 4.1B), suggesting that only the SNP rs4633 is necessary for determining expression rates *in vitro*. Therefore in the case of translation *in vitro*, we can rule out the possibility that the other downstream SNPs of three major COMT haplotypes play a role in determining efficiency of *in vitro* translation.

Computational analysis reveals structural differences

To investigate the structural mechanism in which SNP rs4633 affects *in vitro* translation of COMT RNA, we employed an array of computational approaches. We initially utilize Mfold to generate secondary structures of each haplotype. By examining the folds of COMT mRNA at various sequence lengths, we observe which type of structures predominate at the 5' end. We find that much of the structure toward the 5' end is identical for all haplotypes with the exception near the vicinity of the SNP rs4633 and start codon (Fig. 4.2B). The observation that a single SNP affects the 5' end structure in the vicinity of the start codon supports the view that this region might regulate COMT translation.

LPS and HPS haplotypes exhibit equivalent *in vitro* protein expression levels, which can be attributed to identical secondary structures near the start codon due to sharing the ^{166}C allele. However, the ^{166}T allele structure (carried by the APS haplotype) has some unique structural rearrangements in comparison with the ^{166}C-allelic variant that influences structure stability in the vicinity of the start codon (Fig. 4.2A). Furthermore, consistent with these observations, thermodynamic analysis from OligoWalk suggests that the ^{166}T allele structure is less stable near

the vicinity of the start codon compared to the ^{166}C allele (Fig. 4.2B). Local secondary structures are extremely conserved in the vicinity of the start codon for the ^{166}C allelic variant, exemplified in the optimal and suboptimal structures predicted by Mfold where Stem I is identical for both (Fig. 4.S4). Contrarily, local secondary structures in the vicinity of the start codons differ between optimal and suboptimal structures in T-allele and display more structural diversity.

It is suggestive that the region in the vicinity of the start codon may potentially play a more significant role in regards to translational efficiency (54,67). Since we know that the alternative ^{166}T allele yields a unique structure in the vicinity of the start codon, it is possible to fold this sequence using DMD simulations and seek whether tertiary interactions may play an integral role. We find that both ^{166}C allele and ^{166}T allele predicted structures derived from DMD simulations are identical to secondary structures predicted by Mfold and RNAstructure (Fig. 4.2A; 4.3A,B). However, the native ensemble of ^{166}T allele is less probable in comparison to the ^{166}C allele. The conformational entropy for the ^{166}T allele is higher than the ^{166}C allele, resulting in high flexibility and exploration of higher energy states (Fig. 4.3D). Consequently this may enable facilitated initiation of translation as displayed by the higher expression levels by this haplotype. Our thermodynamic analysis reveals that this is a strong possibility given that the folding transition temperature is lower for ^{166}T allele's structure than ^{166}C.

Codon bias and mRNA structure as factors in translation efficiency

Two main mechanisms are thought to be important factors in determining the efficiency of mRNA translation: 1) the ease of unwinding the mRNA structure at the 5' end and 2) codon usage. The extent to which one mechanism plays a dominating factor over the other is dependent upon the individual genes in question (53,54,74,75). Our simulations in this work

have focused on the structural contributions that can lead to increased protein expression from the ^{166}T allele. Here we consider the possibility of codon bias and its effect on translational regulation of COMT. Both the ^{166}C allele and ^{166}T allele at rs4633 code for histidine through the synonymous codons CAC and CAT, respectively. It has been reported that the CAC allele is nearly twice as efficient compared to the CAT allele (76). Within this context, it is unlikely that the CAT codon is the contributing factor for increased protein expression. However, it has also been reported that use of low efficiency codons near the initiation site can aid the efficiency of translation as this scenario prevents ribosomal traffic jams near the initiation site (77). It is uncertain to what extent the CAT codon increases efficiency of translation given these two competing scenarios. These explanations are in agreement with previously published data on the role of mRNA structure and codon usage in the vicinity of the start codon for translation efficiency (68). Nevertheless, reduced ribosomal traffic may also play a role in the enhancement of ^{166}T allele protein expression.

Role of mRNA stability and dynamics in translation initiation

Previous reports have only suggested a correlation of free energy stability and translational efficiency (41,54). Since it is the stability of the 5' end that is the most determining factor for translation efficiency, it is presumed that the limiting factors are unwinding by the eIF4a factor and ribosomal binding to the start codon. The relationship between free energy and translation initiation is not immediately apparent. The stability of a particular region alone would not necessarily render a recognition site inaccessible. Thus, we proposed that the folding pathway might play an essential role in regulating initiation factor access. Specifically, if the energetic barriers between conformational states of an RNA are low, then the RNA can easily

explore conformations that are outside its native ensemble (73). These conformations can have a reduced number of structural elements and therefore this flexibility can facilitate sequence recognition by translational machinery.

The results presented here support this model. Substitution of ^{166}C to ^{166}T at rs4633 in COMT mRNA increases the number of favorable isoenergetic conformational states for its mRNA transcript. The dynamics of the ^{166}T-carrying allele become entropy driven such that the native conformation becomes less populated as the RNA explores conformations at higher energies. Consequently there are large fluctuations in the positions of each nucleotide and thereby enhancing its flexibility. Further exploration of this model by studying the dynamics for a wide variety of RNA structures would be required to prove its fundamental significance.

Implications for cellular expression levels

Our current results in several transfected mammalian cell lines re-enforce the conclusion from our previous studies (15) that the *in vivo* expression of COMT is strongly dictated by RNA structures formed by SNPs rs4818 and rs4680 (Fig. 4.4) yielding the lowest expression levels for haplotype HPS. However, the *in vitro* translation rate seems to be independent from rs4818 and rs4680 interactions and driven solely by local structures near the start codon that is dependent on allelic variants of rs4633 (Fig. 4.1B). Furthermore, in two out of four cell lines we also observed that the APS haplotype produced the highest protein level, consistent with *in vitro* translation results that are rs4633-dependent. Thus we observed differential input of three SNPs into two distinct structural mechanisms, apparently contributing to translation regulation at different levels.

These results are in line with the observed mRNA structure-dependent differences in efficiency and rate of translation *in vivo* and *in vitro* reported previously (44). Since the rate of translation is much slower than the rate of RNA folding, it is thought that RNA begins to fold locally during the translation process and that the final structure oftentimes is the metastable product of local folding. Thus, the upstream structures dominate folding outcomes *in vitro*, suggesting that folding occurs sequentially. However, when studied *in vivo*, upstream and downstream structures are presented equally and folding outcomes reflect the relative stability of alternative structures, probably facilitated by cellular chaperone proteins associated with nascent RNAs (44).

The variation in COMT expression levels across different systems could potentially be explained by the experimental observation that RNA can adapt specific structures in cells due to rapid exchanging of states facilitated by proteins bound to nascent RNAs (44) in contrast to *in vitro* translation conditions. Because the rate of RNA folding is on the scale of microseconds and thus much faster than the rate of transcription, there is a preference for local folds as opposed to long-ranged base pairs. This preference may be diminished in cells by specific RNA-binding proteins that allow exchange of secondary structures through branch migration (44). Our results suggest that the contribution of these cellular proteins is tissue specific, such that in some cell lines the overall cellular protein expression is almost exclusively controlled by these factors, while other cell lines recapitulate the results found *in vitro* using rabbit reticulocyte lysates (Fig. 4.4).

It is also plausible that in some cell lines, other factors regulating translation are more strongly contributing to protein expression. The abundance of tRNAs with synonymous codons are known to vary in the cell up to ten-fold across different human tissues (78). The availability

of tRNAs during translation could also contribute to the relative speed at which the protein is synthesized.

Alternatively, it is possible that structural modulation of RNA itself is not the sole explanation for differences in protein expression, and there may be additional mechanisms contributing to translational regulation. These downstream structural motifs may potentially be recognized by external biological factors and subject to further regulation. For example, the Fragile X Mental Retardation Protein (FMRP), an established regulator of translation, is known to bind to specific structural RNA motifs (79,80) and can downregulate their expression by association with the RNA-induced silencing complex (81,82). This cascade of structural and cellular mechanisms at the mRNA level is likely to be defined by other specific cellular components and thus contribute to differences in COMT protein expression levels in a tissue-dependent manner.

From a broader perspective, since the APS haplotype carries a non-synonymous met^{158} variation known to create a thermolabile mutant and thus display lower enzymatic activity (12,16) in comparison with wild type val^{158}, it is remarkable that its protein expression level can be significantly higher than wild type LPS haplotype. Thus our results represent a potential compensatory mechanism of APS haplotypes to overcome lower enzymatic activity via overexpression in specific cell lines.

Conclusions

The results presented here demonstrate a new molecular mechanism thereby synonymous substitution of a known functional human COMT haplotype contributes to translation efficiency, thus representing an exciting example of evolutionary selection of an RNA-structure

destabilizing allele to compensate for a destabilizing amino acid substitution within a mutant protein structure. Importantly, this change did not only affect the stability of RNA structure but rather its dynamics, suggesting that increased conformational flexibility enhances translational efficiency. This mechanism by which the destabilizing allele facilitates translation provides a new perspective in functional genomics and requires further investigation to determine the extent of its fundamental applicability for common genetic variations in human population.

FUNDING

This work was supported by the US National Institutes of Health grant [R01GM080742 to N.V.D.]; American Recovery and Reinvestment Act supplements [GM080742-03S1, GM066940-06S1 to N.V.D.]; National Institute of Dental and Craniofacial Research and National Institute of Neurological Disorders and Stroke grants [RO1-DE16558, UO1-DE017018, PO1 NS045685 to L.D.]; Intramural Research Programs of National Center Biotechnology Information a National Library of Medicine to S.A.S.

ACKNOWLEDGEMENTS

We would like to thank Dr. Sergei Romanov for his aid in developing the *in vitro* translation assay.

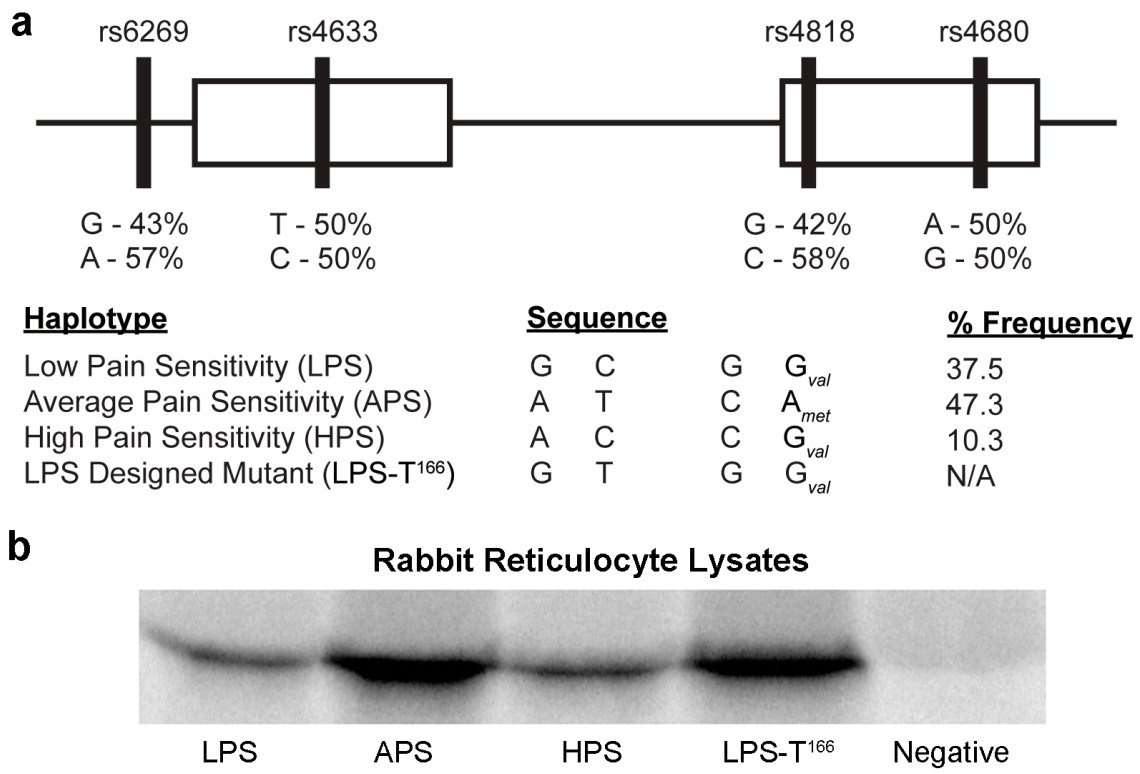


Figure 4.1. Haplotypes of COMT and corresponding expression levels. (**A**) Organization of SNPs for the three haplotypes along COMT gene. Percent frequency of each haplotype from a cohort of 202 healthy Caucasian females is indicated on the right (13). The LPS designed mutant used for later experiments is also included for comparison. (**B**) *In vitro* translation in rabbit reticulocyte lysates. Both LPS and HPS share the same expression levels, while APS exhibits higher expression levels. To test whether this is due to the rs4633 SNP near the start codon, we mutated the LPS variant from ^{166}C to ^{166}T. The resultant LPS-T^{166} mutant displays expression levels on par with APS.

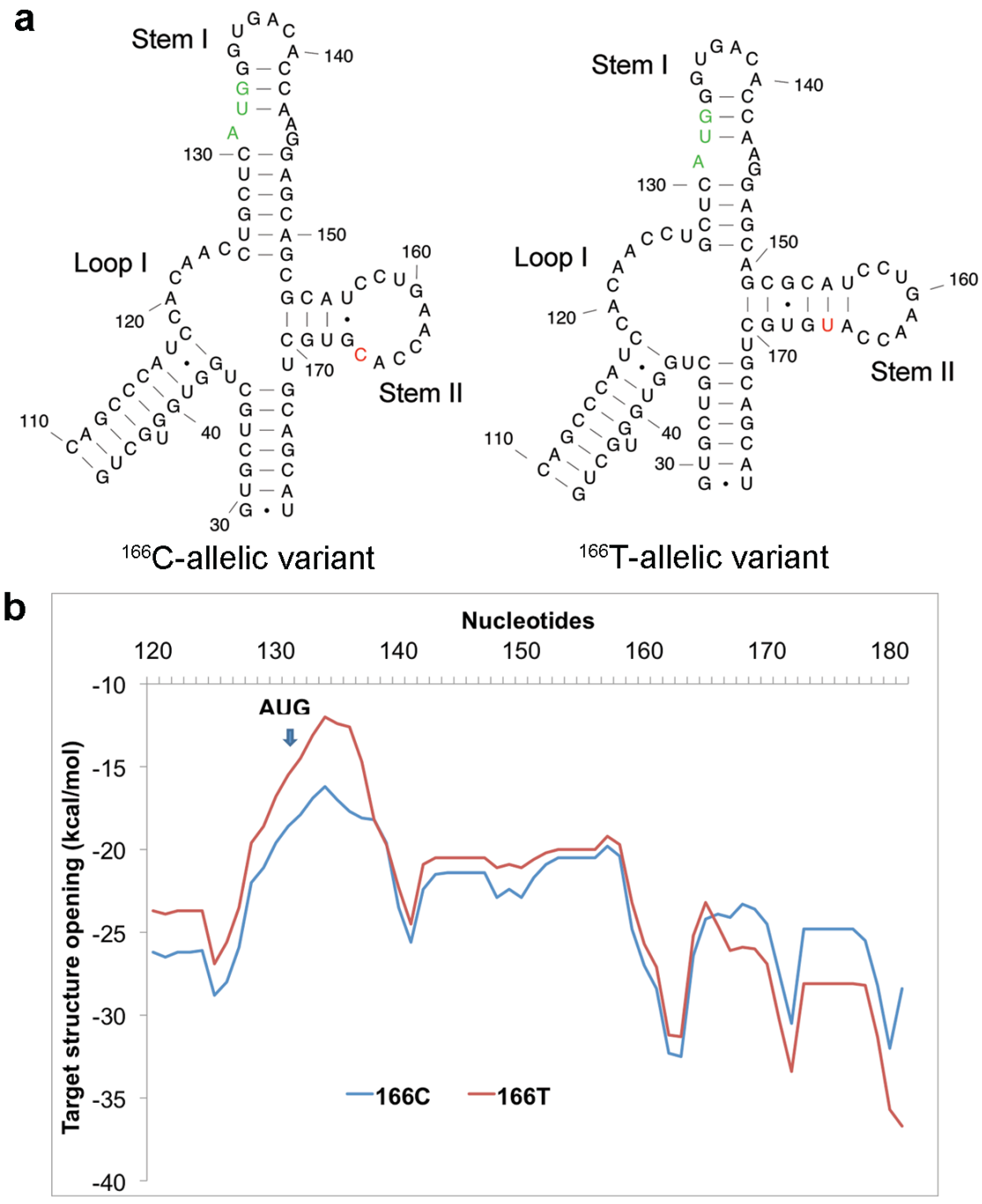


Figure 4.2. Secondary structure analysis of ^{166}C and ^{166}T-allelic variants of COMT by Mfold. (**A**) Free energy profile (ΔG) for target structure opening in the vicinity of the rs4633 SNP and the start codon for ^{166}C and ^{166}T-allelic variants determined using OligoWalk. (**B**) The most probable RNA secondary structural predictions for ^{166}C and ^{166}T-allelic variants near start codon.

We zoom in toward the regional differences between the allelic variants. The start codon is colored in green and the varying SNP is colored in red. The regions of difference are within the motifs labeled Loop I, Stem I and Stem II. The two stems leading to the Loop I and Stem II (capped at base pairs 46-110 and 29-178, respectively) are identical as are the rest of the nucleotides within the 210 nucleotide window (Supplementary Material, Figs. S1-S3). Both DMD and Mfold predict identical secondary structures for the ^{166}C-allelic variant and ^{166}T-allelic variant.

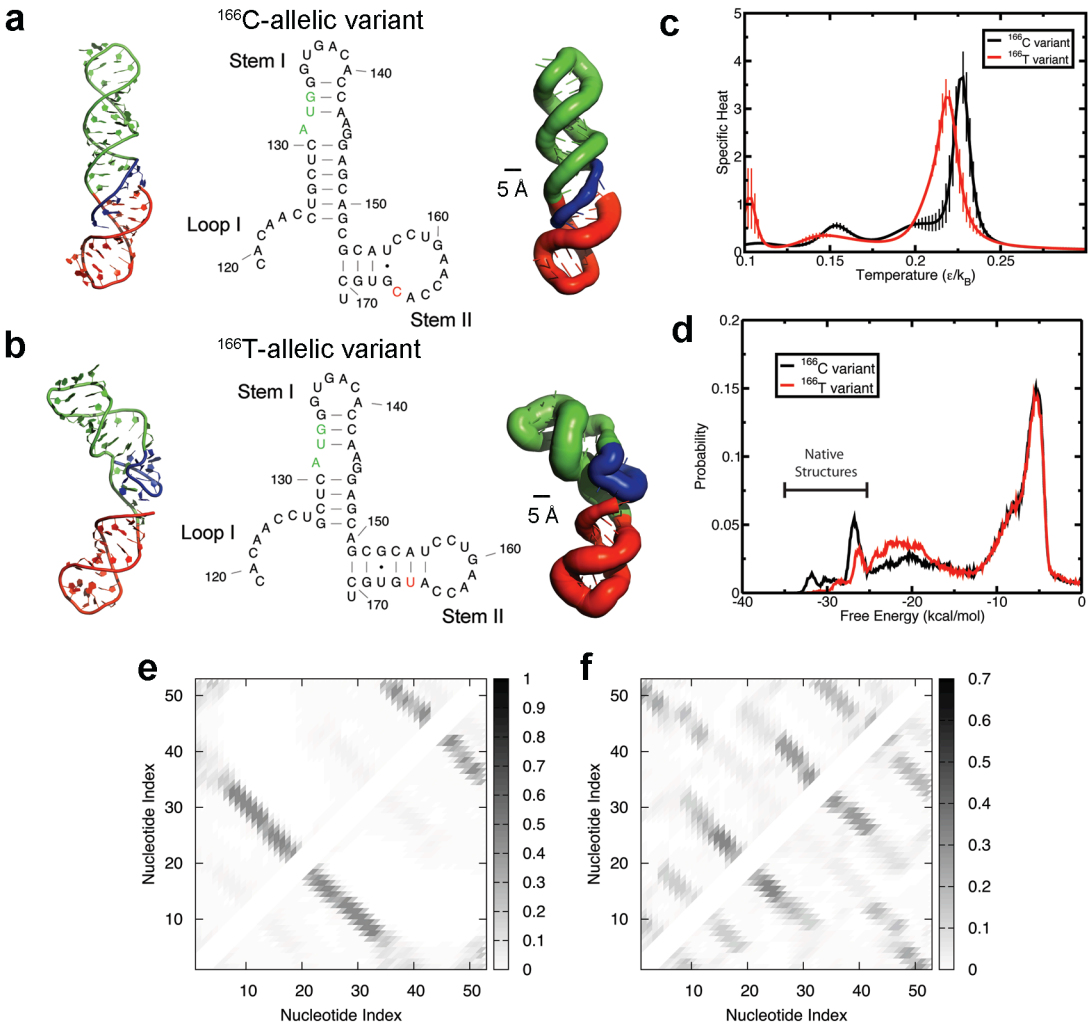


Figure 4.3. Tertiary structure analysis by DMD simulations indicate that the unfolding energy barrier is lower for ^{166}T allele thus enhancing its conformational flexibility to explore higher energy states. (**A** and **B**) Structures of variants as predicted using DMD. Leftmost structures represent the native tertiary structure for each allelic variant. The middle structures are the secondary structural elements as predicted by DMD. The rightmost structures are represented in sausage format where thicker backbones represent larger root mean squared fluctuations within the native ensemble (see **Materials and Methods**). Regions of the RNA are color-coded to correspond to different motifs: Loop I (blue), Stem I (green), and Stem II (red). (**A**) Structures of ^{166}C-allelic variant. (**B**) Structures of ^{166}T-allelic variant. (**C**) The folding transition temperature

for ^{166}T is lower than ^{166}C, as denoted by the peak in specific heat of folding. Specific heats were determined using weighted histogram analysis method. The vertical droplines represent the standard deviation for each plot point. (**D**) Histogram of free energies derived from replica-exchange simulations. The native ensembles that correspond to the structural motifs as predicted by Mfold are denoted on the plot. (**E** and **F**) Maps illustrating frequency of contacts at a temperature of 0.10 ε/k_B. (**E**) Frequency of contacts for ^{166}C-allelic variant. The most frequent contacts shown are representative of the native state alone, where the first lower left diagonal line is representative of Stem I (nucleotides 5 through 32 of the simulation correspond to nucleotides 124 through 151 on the transcript). The second smaller diagonal line on the upper right of the plot represents the contacts formed from the Stem II (nucleotides 34 through 51). (**F**) Frequency of contacts for ^{166}T-allelic variant. Even at the lowest simulation temperature, the ^{166}T SNP creates an ensemble of states where many different enthalpic contacts are made throughout its folding simulation but are entropically unfavorable.

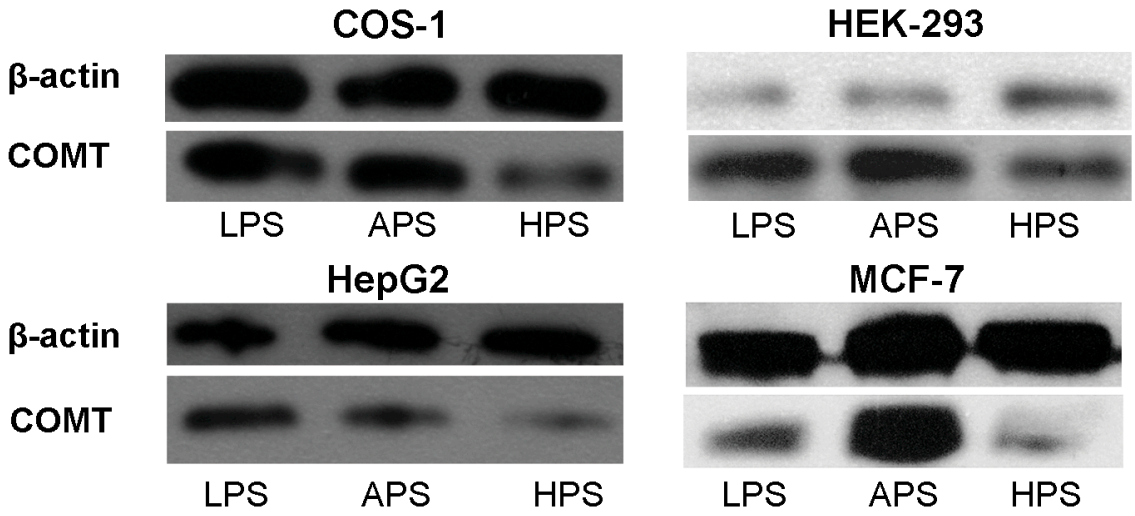


Figure 4.4. Protein expression levels of COMT constructs that code for three COMT haplotypes in the range of mammalian cell lines. The constructs were transiently transfected into indicated cell lines and its protein expression were analyzed via Western blotting. Top band: β-actin. Lower band: COMT. Cell lines: COS-1 monkey kidney cells, HEK-293 human embryonic kidney cells, HepG2 human liver cells, and MCF-7 human breast cells. Among these cell lines, APS protein expression is equivalent or higher than LPS, while HPS exhibits significantly lower levels compared to the two haplotypes.

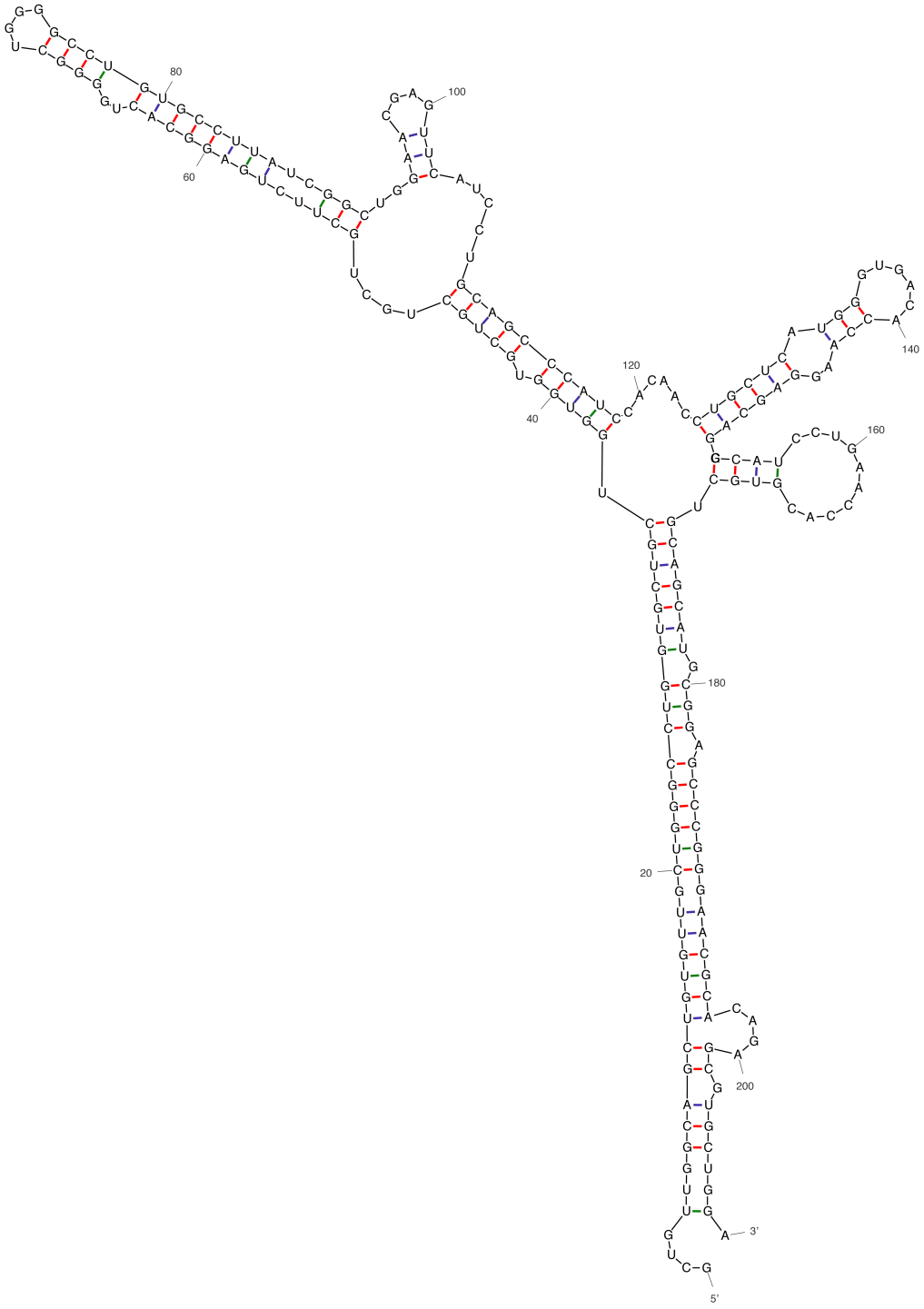


Figure 4.S1. Mfold-predicted structure of LPS from nucleotides 1-210.

Regions different between three haplotypes reside in nucleotides 119-171.

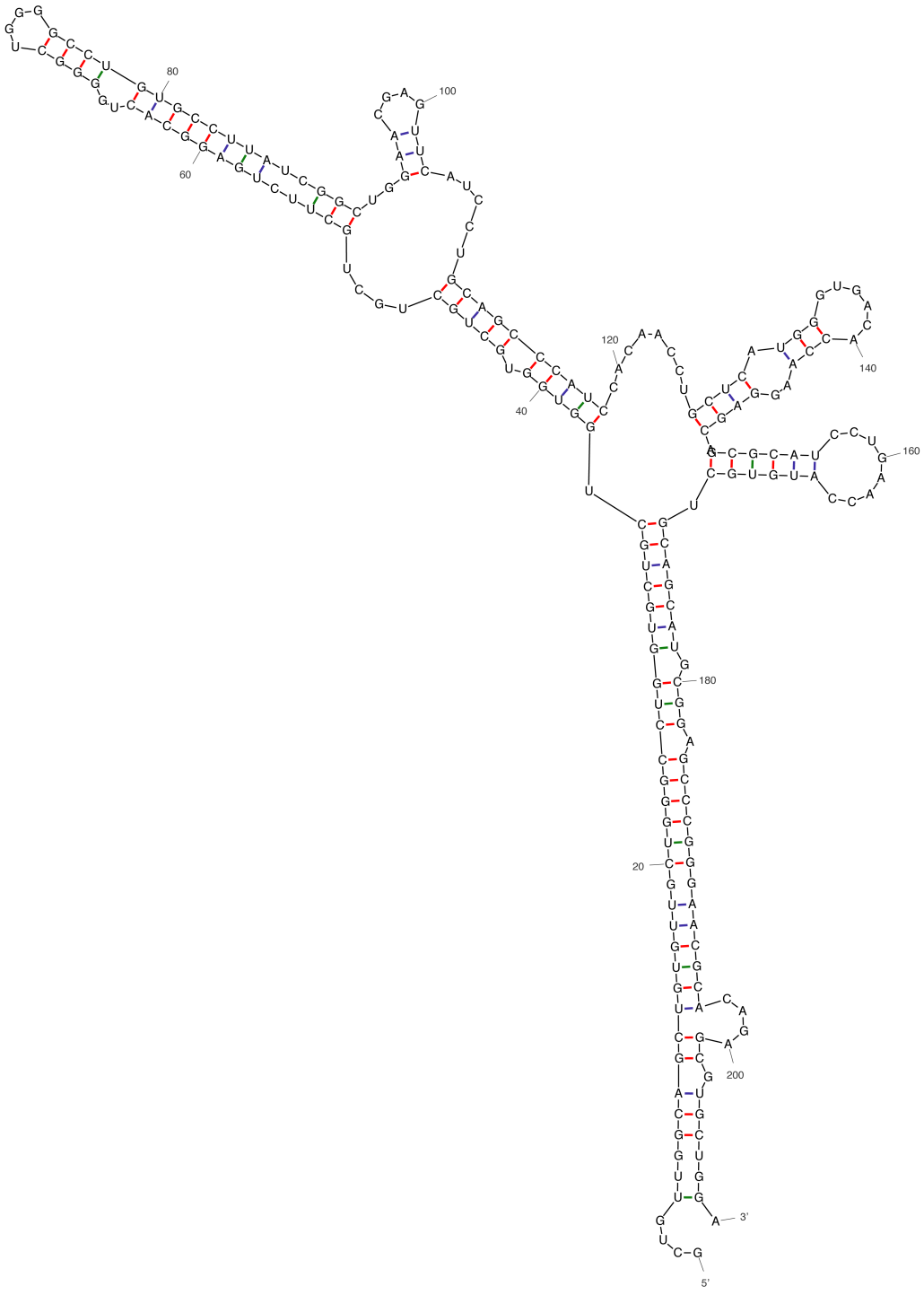


Figure 4.S2. Mfold-predicted structure of APS from nucleotides 1-210.

Regions different between three haplotypes reside in nucleotides 119-171.

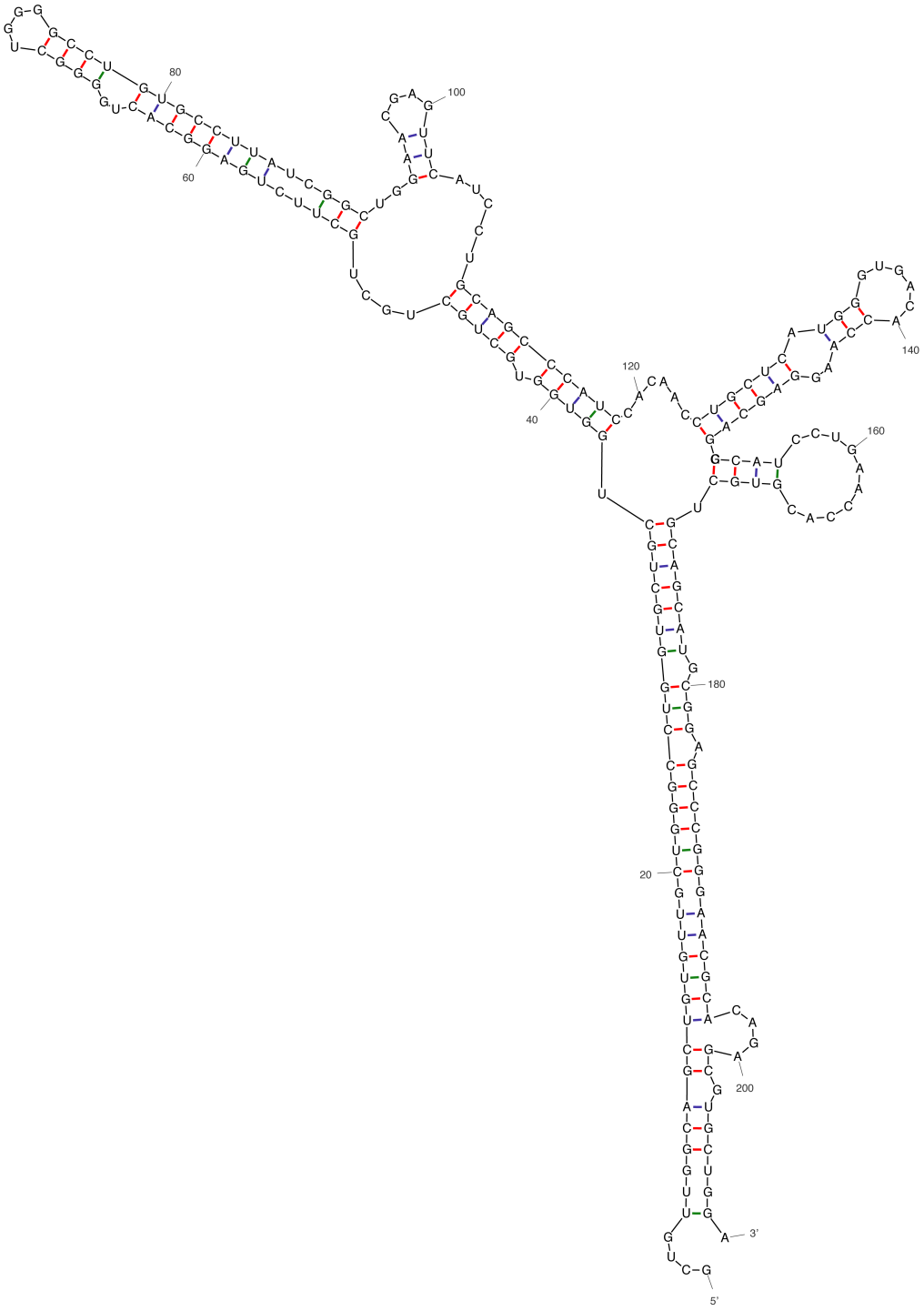


Figure 4.S3. Mfold-predicted structure of HPS from nucleotides 1-210. Regions different between three haplotypes reside in nucleotides 119-171.

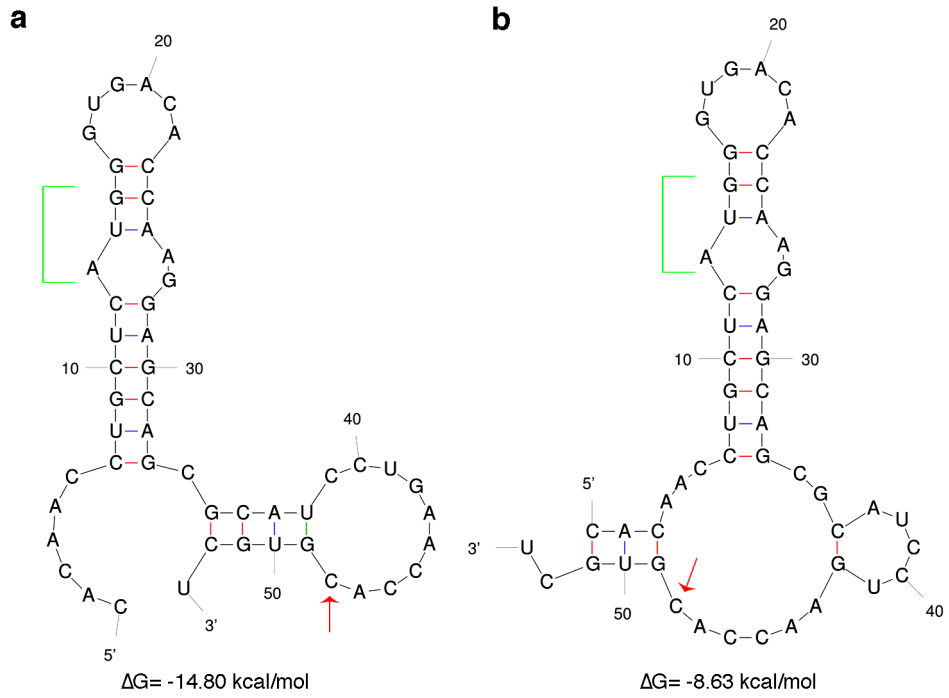


Figure 4.S4. Suboptimal RNA secondary structural prediction for ^{166}C allelic variant near start codon. The structures depicted here are from nucleotides 119 to 171 only focusing regional differences between the allelic variants. Green brackets denote locations of the start codon. Red arrows denote the SNP specific to this allelic variant. (**A**) Optimal structure. (**B**) Suboptimal structure. Only one suboptimal structure for C allelic variant was predicted when percent suboptimality is set to 30 (when only folding within 30% from the minimum free energy will be computed).

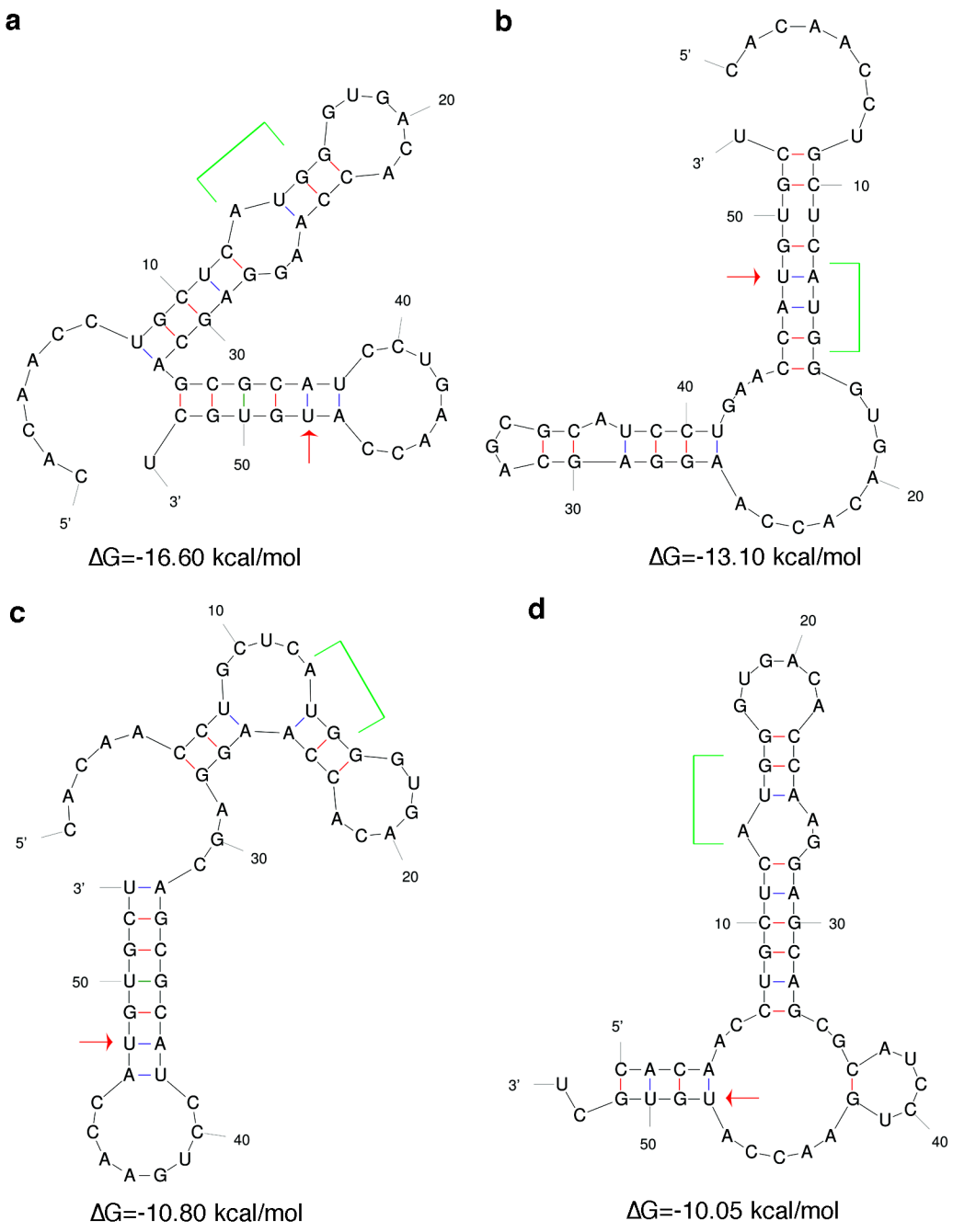


Figure 4.S5. Suboptimal RNA secondary structural predictions for ^{166}T allelic variant near start codon. The structures depicted here are from nucleotides 119 to 171 only focusing regional differences between the allelic variants. Green brackets denote locations of the start codon. Red arrows denote the SNP specific to this allelic variant. (**A**) Optimal structure. (**B-D**) Suboptimal structures. All suboptimal structures for T allelic variant are predicted when percent

suboptimality is set to 30 (when only folding within 30% from the minimum free energy will be computed).

Table 4.S1. Comparison of root mean square fluctuations (RMSF) at each nucleotide for both allelic variants. Values here are used to generate the sausage diagrams as depicted in Fig. 3 of the main text.

Nucleotide #	**^{166}C Variant RMSF**	**^{166}T Variant RMSF**
1	2.20	9.58
2	2.20	9.58
3	1.44	9.77
4	3.21	7.70
5	1.04	7.64
6	1.45	6.64
7	1.95	5.84
8	2.99	6.24
9	1.97	6.77
10	3.24	1.54
11	4.13	3.65
12	2.95	4.73
13	4.57	1.64
14	4.88	4.31
15	3.84	3.35
16	5.48	4.65
17	4.64	7.31
18	4.33	5.51
19	4.68	8.32
20	3.93	10.37
21	3.93	8.92
22	3.80	11.09
23	3.97	11.11
24	3.69	8.92
25	3.77	11.16
26	4.40	9.24
27	4.28	6.84
28	3.96	8.77
29	4.41	4.75

30	3.90	5.74
31	3.91	3.10
32	3.81	5.63
33	3.32	2.23
34	3.22	7.48
35	3.13	7.53
36	2.84	4.55
37	2.78	8.99
38	2.79	8.13
39	2.28	4.67
40	2.80	9.39
41	2.91	8.01
42	2.31	5.96
43	3.43	8.92
44	3.66	6.91
45	3.04	3.59
46	4.40	7.43
47	4.60	3.59
48	3.80	0.51
49	5.43	3.88
50	5.40	4.06
51	4.34	3.12
52	6.32	1.86
53	5.78	2.90

CHAPTER V: SEROTONIN-INDUCED HYPERALGESIA VIA INHIBITION OF CATECHOL O-METHYLTRANSFERASE ACTIVITY

Douglas Tsao[1,2], Samantha Segall[2], Jeffrey Wieskopf[3], Jeff Mogil[3], Nikolay V. Dokholyan[2,4], Luda Diatchenko[2]

[1]Department of Chemistry

University of North Carolina, Chapel Hill, NC 27599, USA;

[2]Center for Neurosensory Disorders, School of Dentistry

University of North Carolina, Chapel Hill, NC 27599, USA;

[3]Department of Psychology, McGill University

1205 Dr. Penfield Avenue, Montreal QC Canada, H3A 1B1;

[4]Department of Biochemistry and Biophysics, School of Medicine

University of North Carolina, Chapel Hill, NC 27599, USA

Corresponding Author: Luda Diatchenko (lbdiatch@email.unc.edu)

ABSTRACT: Peripherally injected serotonin has shown to produce a hyperalgesic effect. Thus far, the 5-HT1A receptor is the major target consistently associated with this response. Here we show that serotonin inhibits catechol O-methyltransferase (COMT), another protein involved in the pain response pathway, via non-competitive inhibition with catechol substrates (K_i = 44 μM). From computational modeling, serotonin actively competes with the methyl donor S-adenosyl-L-methionine (SAM) within the active site. Binding of COMT to

serotonin inhibits the methyl donor from entering the active site, thus preventing methylation of COMT substrates. Animal model studies using mice show that serotonin-induced hyperalgesia is reduced when SAM is co-administered. Our results suggest that inhibition of COMT via serotonin binding contributes to hyperalgesia, and blocking this interaction with SAM may provide another strategy to potentiate analgesic efficacy of serotonin; for example, by selective serotonin reuptake inhibitor treatments.

INTRODUCTION

Anti-nociception associated with serotonin (5-hydroxytryptamine, or 5-HT) has been a continued area of study for the past several decades (83,84). Drugs known to act on 5-HT receptors have been shown to lower the response to noxious stimuli when administered (85). Serotonin concentration is regulated in patients by prescribing selective serotonin reuptake inhibitors (SSRIs) (86). This class of compounds acts to increase the extracellular concentration of serotonin by preventing the presynaptic cell from its reuptake and therefore allowing the postsynaptic receptors to continue with serotonin binding. Most commonly, serotonin concentrations are regulated in patients to treat for disorders such as depression and anxiety. SSRIs are part of a much larger family of anti-depressant compounds.

Anti-depressants are also used in the clinic to treat chronic pain conditions, such as fibromyalgia and rheumatoid arthritis (86). However, the mechanisms responsible for these analgesic effects are poorly understood (87). It is known that anti-depressants can differ in their analgesic efficacy (87). SSRIs exhibit lower analgesic properties in comparison to other anti-

depressants. Similarly, injecting serotonin peripherally to the brain is less efficacious as opposed to increasing serotonin concentration indirectly by administering SSRIs (88).

Furthermore, analgesia is not an exclusive property of serotonin. Serotonin is also known to produce a hyperalgesic response when injected peripherally to other tissues (89). A specific class of 5-HT receptors (5-HT1A) is believed to be responsible for creating a hyperalgesic response; however, the extent it contributes is unknown. The perception of pain is a complex response in which several classes of proteins are involved, and the extent that they are implicated is dependent upon the received noxious stimulus.

Catechol O-methyltransferase (COMT) is an enzyme that has been implicated in the perception of mechanical, thermal, ischemic, and inflammatory pain in both humans and rodents (13,17,90). Its association with broad range of noxious stimuli suggests its critical role as an underlying factor for the fundamental processes of pain perception. The enzymatic activity of COMT is correlated with the sense of perceived pain; higher levels of COMT activity correspond to lower pain sensitivity (13,15,17,90). Thus, compounds or cellular regulatory factors that lower COMT activity can enhance the pain sensation in individuals.

A co-factor necessary for methylation of COMT substrates is S-adenosyl-L-methionine (SAM). Comparing the chemical structures of SAMs and serotonin, we see that the indole ring of serotonin is similar to the adenosine motif of SAM. Thus, we hypothesize that serotonin may play a role as a competitive inhibitor of SAMe and thus provide an avenue for COMT-dependent pathway in serotonin-induced hyperalgesia. Specifically, the inhibition of COMT activity through serotonin would reduce the ability of COMT to methylate its native catecholamine substrates, including several neurotransmitters, such as dopamine, epinephrine, and norepinephrine (1). Preventing methylation of epinephrine and norepinephrine will lead to

continuous activation of a pain signaling through the β_2 – and β_3–adrenergic receptor pathways(17). Here, we investigate the binding between serotonin and COMT using a combination of *in silico*, *in vitro*, and animal model studies.

MATERIALS AND METHODS

Docking Simulations of Serotonin to COMT Active Site

We employ the MedusaDock package to generate possible ligand conformations within the protein active site, and utilize the MedusaScore package to evaluate each conformation generated by MedusaDock (26,27). MedusaDock enables flexible docking of both the ligand and the side-chain amino acids of the protein. We perform all ligand docking simulations using the crystal structure of human COMT (PDB: 3BWM) (9).

Prior to all docking simulations, we first stripped all crystallized ligands bound to COMT. Crystallographic waters that were in the active site were retained for our docking simulations. We performed 200 docking simulations for serotonin, where each simulation began with a different seed number and each conformation generated from a simulation was subsequently minimized. All conformations were then ranked according to their free energy value. The lowest energy structure is determined to be the native pose.

Binding Measurements using Surface Plasmon Resonance

Purified COMT protein (derived from porcine liver) and serotonin ligand was obtained from Sigma-Aldrich. Protein concentration was verified by performing a BCA assay (Thermo Scientific). Purified COMT was biotinylated at its surface lysine residues using EZ-Link NHS-

LC-LC Biotin (Thermo Scientific) at a 1:1 mole ratio of biotinylating reagent to protein and subsequently purified using a 1 mL Sephadex G-25 medium spin column. Biotinylated COMT was kept in a buffer (10 mM Tris (pH 7.4), 1 mM $MgCl_2$, and 1 μM DTT) and stored at 4 °C until use.

Surface plasmon resonance (SPR) measurements were conducted at 25 °C using a Biacore 2000 instrument with a previously established protocol (91). Biotinylated COMT was first attached to a streptavidin-coated flow cell (sensor chip SA or Biotin CAPture Kit, GE Healthcare), followed by a buffer flow, introduction of serotonin-containing buffer, and then a final wash with buffer. Here the flow rate was established at 5 μL/min.

The dissociation rate of the serotonin-COMT complex was determined by fitting the dissociation data to a single-exponential decay.

$$R = R_0 \times e^{-k_d(t-t_0)}$$

Here, the response unit from the SPR machine is denoted as *R*. Once the dissociation rate is determined, the association rate of serotonin to COMT can be found by fitting to the following equation:

$$R = R_{eq}\left[1 - e^{-(k_a[\text{Serotonin}]+k_d)(t-t_0)}\right]$$

Radiolabeled SAM Activity Assay

Activity of COMT was measured through a series of experiments where the enzyme was incubated with the substrate catechol (dissolved in H_2O, Fisher Scientific) and S-[methyl-^{3}H]-adenosyl-L-methionine (PerkinElmer, 10.0 Ci/mmol) (92). Reactions were carried out in a buffer containing 10 mM Tris (pH 7.4), 1 mM $MgCl_2$, and 1 μM DTT with a total reaction volume of 100 μL.

To determine the linear range of the activity assay, SAM concentration (1 μM) was held constant while COMT concentration was varied (10 – 80 ng/μL). Each reaction was placed in a clean PCR tube and incubated at 37 °C for 30 minutes. After incubation, reactions were terminated by addition of 100 μL of 1M HCl to each tube. Radiolabeled catechol products were extracted from the reaction by adding 10 mL of scintillation fluid (MonoFlow I, National Diagnostics) and quantified using a scintillation counter. Activities were normalized for each reaction by performing a duplicate reaction with the COMT inhibitor, OR-486 (0.5 mg of inhibitor dissolved in 30 μL DMSO and 20 μL of 0.9% saline; OR-486 buffer alone does not inhibit reaction), and subtracting its radioactivity. We found that the linear range was between 10 and 40 ng/μL, and thus for all subsequent experiments we used a COMT concentration of 20 ng/μL.

Kinetic experiments were performed at a variety of catechol concentrations (0, 0.5, 1, 3.3, 5, 10, and 33 μM) while maintaining constant SAM (1 μM) and COMT (20 ng/μL) concentrations. Reactions were set-up for a variety of time points: 0, 5, 10, 30, 60, and 120 minutes. Each time-point reaction is incubated at 37 °C and then quenched with 1M HCl at its specified time point. Products were quantified as discussed above. Serotonin-inhibition experiments were conducted for the same catechol concentrations, but with the addition of 100 μM of serotonin (Sigma) to each reaction.

Kinetic parameters were determined by first fitting the data sets to a single-site Michaelis-Menten model (for reference), followed by a two-site Michaelis-Menten model as given by (12):

$$v = V_{\max} \frac{[\mathrm{SAM}][\mathrm{Cat}]}{\left([\mathrm{SAM}][\mathrm{Cat}] + K_{m,\mathrm{SAM}}[\mathrm{Cat}] + K_{m,\mathrm{Cat}} + K_{D,\mathrm{SAM}}K_{m,\mathrm{Cat}}\right)}$$

where catechol is abbreviated as "Cat." To determine the mechanism of serotonin-inhibition, we calculated the apparent K_i ($K_{i,app}$) of inhibition for both a competitive and non-competitive inhibitor, and compared which $K_{i,app}$ yielded the lowest value. The apparent K_i for a competitive inhibitor is given by

$$K_{i,app} = \frac{K_{m,\mathrm{Cat}}[\mathrm{Serotonin}]}{K_{m,app} - K_{m,\mathrm{Cat}}}$$

where $K_{m,app}$ is the apparent $K_{m,\mathrm{Cat}}$ for the serotonin-inhibited curve. The apparent K_i for a noncompetitive inhibitor is given by

$$K_{i,app} = \frac{V_{\mathrm{max,app}}[\mathrm{Serotonin}]}{V_{\mathrm{max}}}$$

where $V_{max,app}$ is the apparent V_{max} for the serotonin-inhibited curve.

von Frey Experiments with Mice

For the automated von Frey tests, mice were placed in individual transparent Plexiglas cubicles set atop a perforated metal floor (with 5-mm diameter holes 7 mm apart), and separated from each other by opaque dividers. The automated von Frey fiber (Ugo Basile Dynamic Plantar Aesthesiometer) was applied to the mid-plantar hind paw. Prior to testing, the mice were habituated to their surroundings for one hour. Three separate threshold determinations were taken at baseline and then averaged. Following the injections, one threshold determination was made on each hind paw at every time point.

RESULTS AND DISCUSSION

Serotonin Binds to COMT

Binding interactions between COMT and serotonin were verified through surface plasmon resonance (Figure 5.1). From the data, both the rate of association (k_{on}) and dissociation (k_{off}) can be determined. We find that serotonin dissociates from COMT with a k_{off} rate of 0.0012 s^{-1}, and binds to COMT with a k_{on} rate of 13.7404 $M^{-1}s^{-1}$. The equilibrium dissociation constant (K_D) was determined by dividing the dissociation rate from the association rate, and calculated to be 87 μM. A calculated 87 μM K_D for serotonin is a reasonable binding affinity when comparing to previously reported values for COMT ligands (12). The lowest reported K_D affinity reported among COMT ligands was for SAM, with a K_D of 20.2 μM. In contrast, many common COMT substrates exhibited dissociation constants in the range of 40 – 400 μM. The higher dissociation constants observed for COMT substrates is most likely a consequence of its low substrate specificity. The binding site for catechol is quite solvent accessible and thus can accommodate many ligands (9). Although COMT will not methylate most phenols or gallols (2), virtually any catechol is susceptible for methylation including bulkier steroid compounds. Thus, with a K_D among the same order of magnitude in comparison to existing substrates, the interaction between serotonin and COMT could be viable within the cell.

Computational Docking of Serotonin to COMT

Next we investigated how serotonin would be accommodated within the active site of COMT. Using computational docking tools, we dock serotonin to an apo-COMT protein to

predict its native pose in the active site. We find that serotonin occupies a site in the pocket that sterically overlaps with the binding site of SAM (Figure 5.2A). In the crystal structure of holo-COMT, His142 participates in an edge-to-face interaction with the adenosine of SAM (9). However, during serotonin binding, His142 donates a hydrogen bond with the amine sidechain of serotonin (Figure 5.2B). Another hydrogen bonding interaction takes place with the serotonin amine and the sidechain carboxyl group of Asp90. These two residues are responsible for positioning serotonin near the binding pocket of SAM, thereby preventing SAM from accessing its site. The carbonyl backbone of His142 also accepts a hydrogen bond from the hydroxyl group of serotonin.

Two more residues form hydrogen-bonding interactions with serotonin. The carbonyl of Met40 forms a hydrogen bond with the imidazole group of serotonin, while Lys144 forms a hydrogen bond with the hydroxyl group of serotonin. The sidechain of Met40 also packs against the indole ring of serotonin. Overall, the results of this docking simulation suggest that serotonin may inhibit COMT activity by preventing the cofactor SAM from binding.

Serotonin Inhibition of COMT Activity

To test whether the binding of serotonin leads to inhibition of COMT activity, we performed a set of kinetics experiments to measure the differences of reaction rates both in the presence and absence of serotonin (Figure 5.3A,B). We measured the variance in reaction velocity with respect to changes in catechol concentration. A double-reciprocal plot comparing COMT velocities in the presence and absence of serotonin shows that addition of 100 μM of serotonin decreases the reaction velocity of COMT (Figure 5.3C). Thus, serotonin binds COMT to inhibit its activity, rather than activate COMT. These results are in line with our

computational model, since serotonin occupies a region within the active site that may be hindering SAM binding. Furthermore, serotonin does not possess a methyl group to donate which would have compensated for the lack of SAM binding.

We determined the mechanism by which serotonin inhibits COMT by fitting our data to a two-site Michaelis-Menten model (see **Materials and Methods**; Figure 5.3D). The Michaelis-Menten equilibrium constant for catechol binding was determined for both in the absence of serotonin (K_M) and in the presence of 100 μM serotonin ($K_{M,app}$). A comparison of K_M (16 μM) and $K_{M,app}$ (18.5 μM) shows a negligible difference between their affinities toward catechol. However, a comparison of the maximum velocity rates reveals a greater difference. The V_{max} in the absence of serotonin was 16 nM/min while the apparent $V_{max,app}$ in the presence of serotonin was 7 nM/min. Our data suggests that serotonin is a non-competitive inhibitor with respect to catechol, since the inhibitory effect is found in the maximum reaction rate as opposed to the binding affinity for catechol. This mechanism is in line with our structural model, since the site for serotonin binding is located near the SAM binding site, and would still allow for catechol binding from a steric viewpoint.

Given the differences between V_{max} and $V_{max,app}$, we can calculate the inhibition constant for serotonin ($K_{i,app}$). We find that $K_{i,app}$ is 44 μM, which is within the vicinity of our approximate K_D determined *via* SPR (~50 μM). Overall, our results show that serotonin binds to COMT and inhibits its activity at a concentration that is similar to natural COMT substrates.

Serotonin-Induced Hyperalgesia Rescued by Addition of SAM in Mice

Our *in vitro* results and *in silico* models suggest that serotonin is an inhibitor COMT. However, whether this inhibition substantially contributes to the observed serotonin-induced

hyperalgesia cannot be determined from this data alone. Our structural model suggests that serotonin inhibits COMT activity by actively competing with SAM at the active site. This mechanism is further supported by our kinetics studies. Thus, by increasing the concentration of SAM we can lower the probability of serotonin occupying the active site of COMT.

If serotonin inhibition of COMT through competitive inhibition with SAM contributes to hyperalgesia, we would expect that administration of SAM would lower the observed hyperalgesia induced by serotonin. We tested this hypothesis by performing several withdrawal threshold experiments with intraplantar administration of serotonin to mice (93).

Mice were fed a diet that included 10mg/kg of SAM per day for a total of 10 days. Repetition of the withdrawal threshold assay with this set of mice (n = 4) showed a reduced hyperalgesic response to serotonin. The results of our mouse models indicate that serotonin-induced hyperalgesia is a response that is associated with COMT inhibition. Furthermore, our administration of SAM indicates that serotonin is indeed actively competing with SAM for the active site of COMT.

Implications for Serotonin-Induced Hyperalgesia and SSRIs

Our current results, consistent with a line of previously published observation(93), show that there is a reproducible effect of hyperalgesia associated with serotonin. Yet, increasing the concentration of serotonin is also commonly used as an analgesic strategy for clinical treatment of chronic pain conditions (86,94). Selective serotonin reuptake inhibitors (SSRIs) act to increase the concentration of serotonin by inhibiting its reuptake by the presynaptic, and are administered for conditions such as fibromyalgia, rheumatoid arthritis, and migraines (86).

However, SSRIs are most commonly prescribed as an anti-depressant. Hence, there is a clear dichotomy in serotonin contribution to pain states.

The pain response can be attributed to many factors, including nociception and apperception. Anti-depressants are commonly prescribed to aid chronic pain patients as an analgesic, although their mechanisms of action are not fully understood (87). The analgesic effects of SSRIs are very modest. A possible explanation for their diminished efficacy may be due to a simultaneous hyperalgesic effect caused by the increase in serotonin concentration and, based on our results, this hyperalgesic effect can be at least in part attributed to serotonin-dependent COMT inhibition. Thus, our results open up a new avenue for increasing the analgesic efficacy of SSRIs by co-administrating with SAM.

CONCLUSIONS

The hyperalgesic effect associated with serotonin is likely a response attributed to a variety of proteins. While the 5-HT1A receptor has been known to produce nociceptive effects, we have shown that COMT is another protein that can cause a hyperalgesic response in the presence of serotonin. The mechanism behind COMT-inhibited hyperalgesia is likely due to the competition between SAM and serotonin for the active site. By further understanding the mechanisms behind serotonin-induced hyperalgesia, co-administrating the appropriate inhibitors or cofactors could increase its analgesic efficacy.

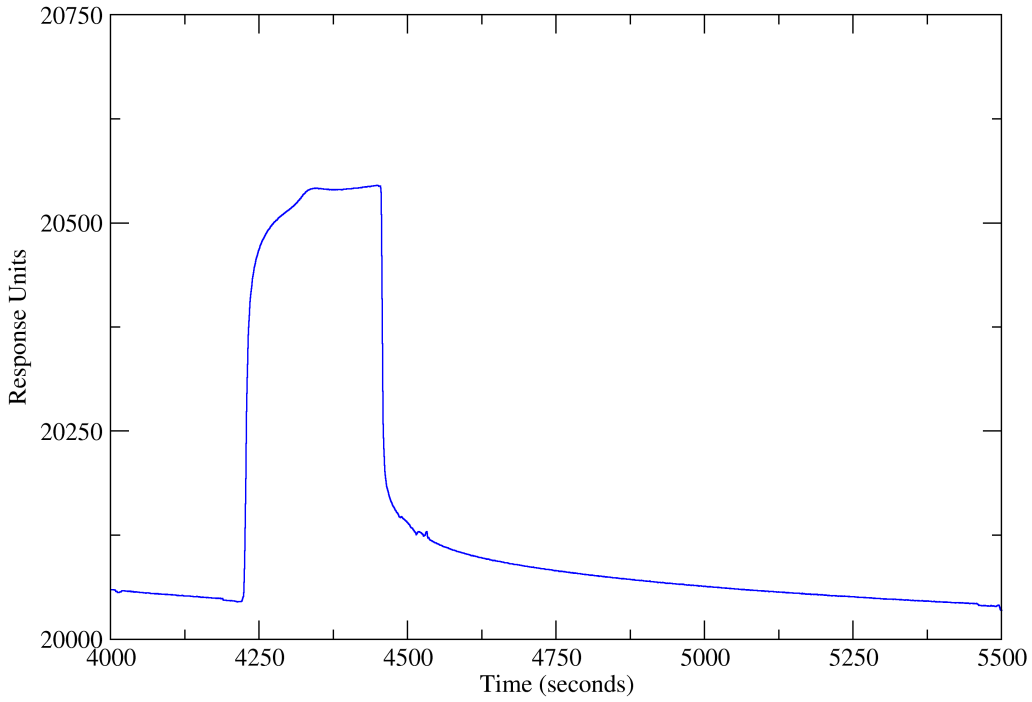


Figure 5.1. Serotonin binds to COMT. SPR data is fitted to association and dissociation curves to determine the kinetics of serotonin binding (k_{on} = 0.0012 s^{-1}, k_{off} = 13.7404 $M^{-1}s^{-1}$).

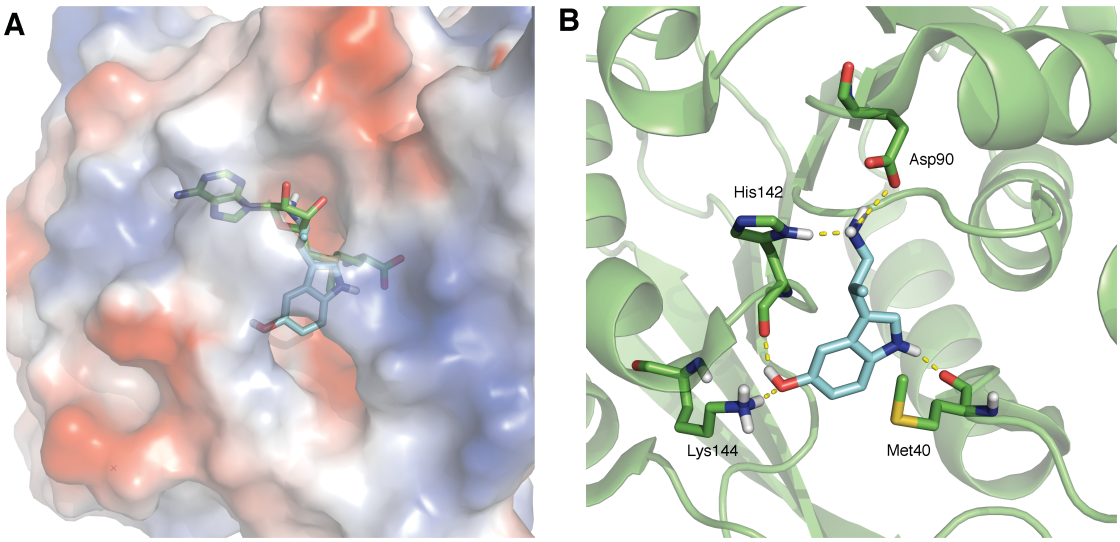


Figure 5.2. Structural model of serotonin within COMT active site. (**A**) Surface representation of serotonin binding to a pocket within COMT. Serotonin structure is shown in cyan; SAM structure is shown in green. The overlapped structures indicate that the amine sidechain of serotonin prevents SAM from actually binding to COMT. On the COMT surface, blue represents positively-charged, red represents negatively-charged, and white represents neutrally-charged regions. (**B**) Interactions of serotonin with residues inside COMT active site. COMT structure and residues are shown in green; serotonin is shown in blue. Hydrogen-bonding interactions are highlighted with yellow-dashed lines.

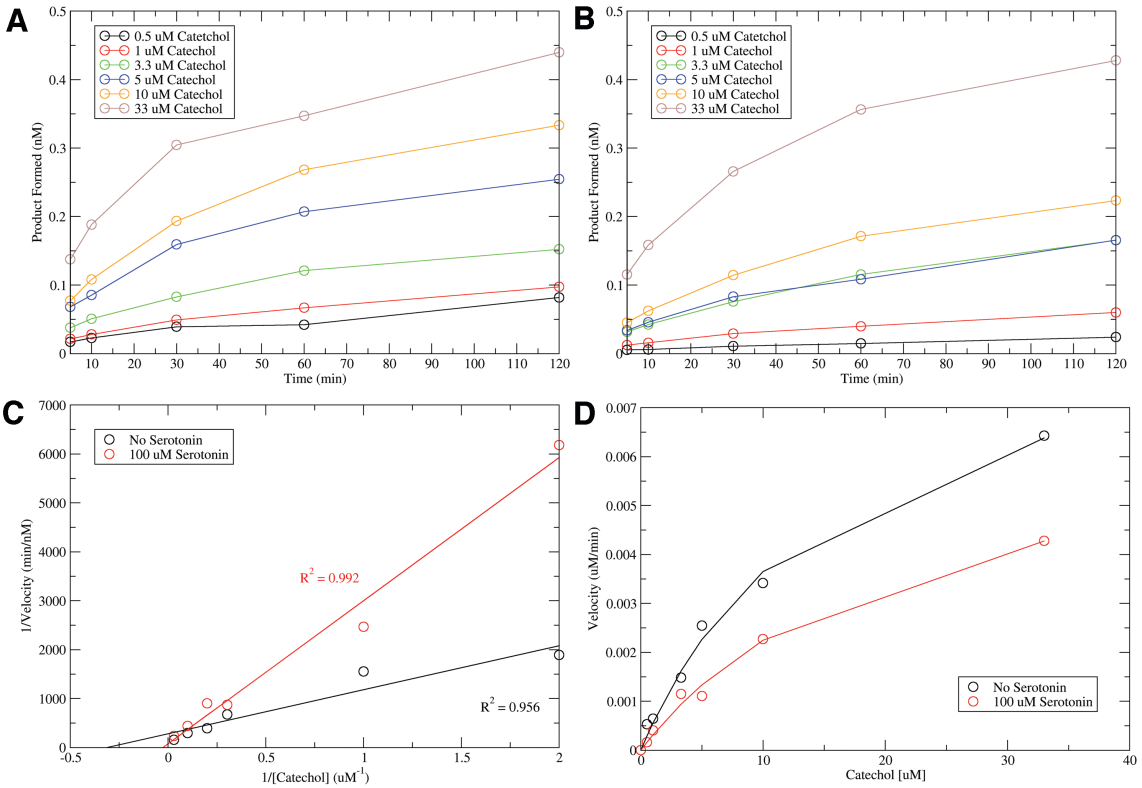


Figure 5.3. Inhibition of COMT with serotonin. (**A**) and (**B**) show the reaction rates of COMT at different catechol concentrations in the absence and presence of 100 μM serotonin, respectively. (**C**) Lineweaver-Burk plot comparing the reciprocal velocities of activity against the reciprocal concentration of catechol in the absence (black) and presence (red) of 100 μM serotonin. The higher slope observed for serotonin indicates some level of inhibition. (**D**) Michaelis-Menten determination of equilibrium inhibition constant. Activity assays in the absence (black) and presence (red) of 100 μM serotonin were fit to a two-site Michaelis-Menten model. While the K_M for catechol remains relatively the same, the V_{max} shows some noticeable differences thereby suggesting a non-competitive mechanism of inhibition between serotonin and catechol.

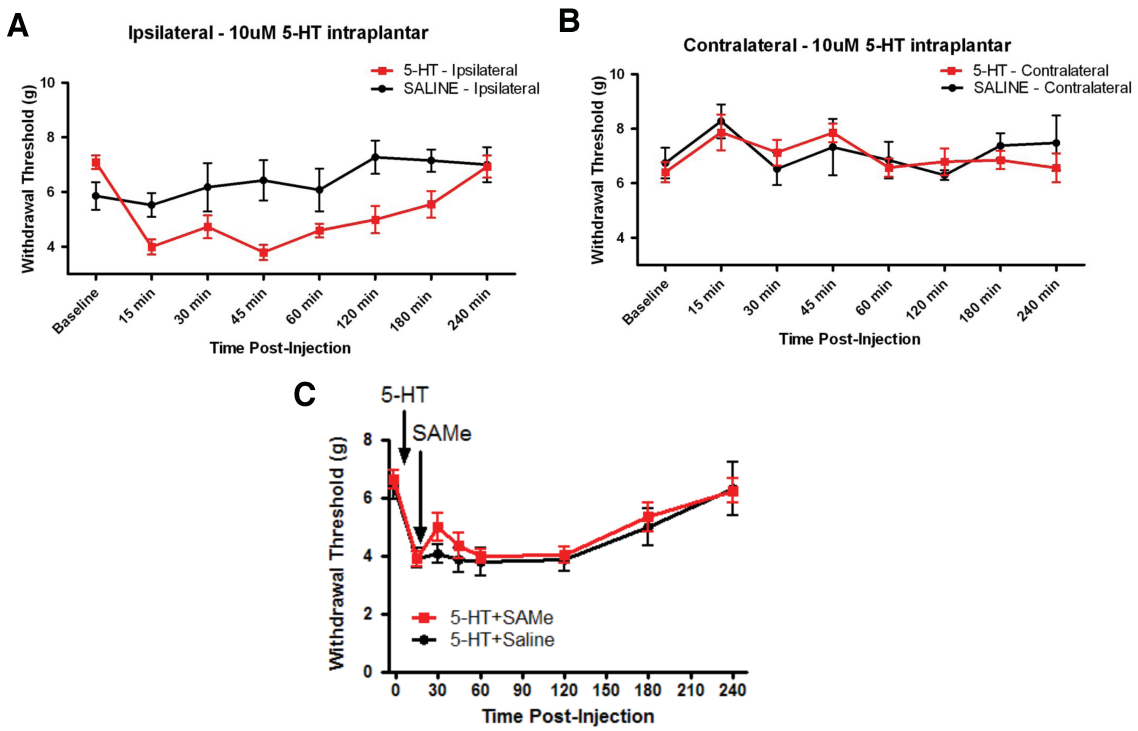


Figure 5.4. Mouse models of serotonin-induced hyperalgesia. (**A**) and (**B**) Serotonin was injected into the ipsilateral and contralateral paw, respectively. We find that the withdrawal threshold for the paws with serotonin (red) injected were statistically smaller compared to those with saline (black). (**C**) Demonstration that increasing SAM concentration lowers serotonin-induced hyperalgesia. Mice with a diet including 10 mg/kg of SAM per day (red) showed a higher withdrawal threshold when administered serotonin compared to mice without a regular diet of SAM (black).

CHAPTER VI: CONCLUDING REMARKS

We have discussed several structural features of COMT that leads to further understanding of its enzymatic activity. Crystal structures of COMT have confirmed a kinetic mechanism of how its ligands are accommodated within its active site: i.e., SAM is initially bound, followed by a divalent metal cation and then the substrate. Previous biochemical data confirmed that a metal is necessary for catalysis; here, we proposed a mechanism that the metal is necessary not only for coordination of the catechol, but to correctly align the sulfonium group as well. Our docking studies discussed here leads to a further understanding of the fundamental mechanism by which COMT conducts catalysis.

Regioselectivity has been an overlooked feature of COMT until now. It is interesting to note that this feature has been evolutionarily conserved throughout nature. Our quantum calculations produce reaction barrier height differences between the two possible methylation positions of each catechol, revealing that the regioselectivity of COMT enhances its catalytic rate. To our knowledge, this is the first example of regioselectivity as a means of catalytic enhancement as opposed to chemical specificity. Thus, our finding here not only broadens our knowledge of how COMT lowers the barrier for its methylation reaction, but also proposes a potentially significant alternative role for regioselection in other enzymes.

The functionality of COMT is not limited to its structural features, but also how it is regulated within the cell. Metabolites within the celluar cytoplasm are often able to affect the functionality of various enzymes *via* means such as allosteric activation or direct inhibition. Here, we have discussed an example of COMT regulation through serotonin inhibition. Serotonin is a monoamine neurotransmitter that is synthesized from the amino acid tryptophan

within the body. The presence of serotonin is known to produce both algesic and analgesic effects. We find that a reason why SSRIs may have such modest analgesic effects compared to other anti-depressants is due to its inhibition of COMT. Specifically, we observed that serotonin is inhibiting COMT activity by competing with the methyl donor SAM for the active site. Thus, a clinically relevant solution to increase SSRI analgesia is to co-administrate with SAM, as we observed for our mouse models.

The regulation of COMT protein expression is another method for the cell to manage COMT enzymatic activity. We find that COMT protein expression is regulated in some cell lines via the translational efficiency of its mRNA. Specifically, we observed that the APS haplotype that encodes for a mutant isoform of COMT ($Val^{108}Met$) contains a SNP within the 5' end of its mRNA that enables higher translational efficiency compared to the other haplotypes. Our secondary and tertiary structural analysis of COMT mRNA for each haplotype shows that the dynamics at the 5' end vary depending on the SNP present. The C→T mutation found at position 166 of the mRNA strand coding for APS creates a highly flexible and partially folded region that facilitates translation for the ribosome. Our work on COMT mRNA regulation contains two main conclusions. First, the allelic variant responsible for increased translation efficiency is located exclusively on the haplotype coding for the low activity protein, suggesting evolutionary selection of an RNA-structure destabilizing allele to compensate for the low activity mutation present within its protein structure. Second, previous works have shown only a correlation between the free energy of the 5' end of a RNA transcript and protein expression. Here we provide a mechanism whereby destabilizing alleles may facilitate translation initiation. These findings have broad biological implications and suggest potential applicability for other common genetic variations in the human population.

Our structural studies of COMT have led insight into both its enzymatic activity and potential mechanisms of regulation. Furthermore, by exploring the properties of COMT we have uncovered additional methods by which other enzymes may function. For example, regioselection may be a property that other enzymes possess to lower the activation barrier for their respective catalyzed reactions. Likewise, genes may evolve to increase its protein expression by selecting for nucleotides that disable its ability to form structurally stable mRNA structures. These mechanisms provide a blueprint toward identifying additional modes of activity and regulation for other enzymes.

REFERENCES

1. Mannisto, P.T. and Kaakkola, S. (1999) Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol Rev*, **51**, 593-628.

2. Axelrod, J. and Tomchick, R. (1958) Enzymatic O-methylation of epinephrine and other catechols. *J Biol Chem*, **233**, 702-705.

3. Roca, M., Andres, J., Moliner, V., Tunon, I. and Bertran, J. (2005) On the nature of the transition state in catechol O-methyltransferase. A complementary study based on molecular dynamics and potential energy surface explorations. *J. Am. Chem. Soc.*, **127**, 10648-10655.

4. Tilgmann, C., Melen, K., Lundstrom, K., Jalanko, A., Julkunen, I., Kalkkinen, N. and Ulmanen, I. (1992) Expression of recombinant soluble and membrane-bound catechol O-methyltransferase in eukaryotic cells and identification of the respective enzymes in rat brain. *Eur J Biochem*, **207**, 813-821.

5. Ulmanen, I., Peranen, J., Tenhunen, J., Tilgmann, C., Karhunen, T., Panula, P., Bernasconi, L., Aubry, J.P. and Lundstrom, K. (1997) Expression and intracellular localization of catechol O-methyltransferase in transfected mammalian cells. *Eur J Biochem*, **243**, 452-459.

6. Tenhunen, J., Salminen, M., Lundstrom, K., Kiviluoto, T., Savolainen, R. and Ulmanen, I. (1994) Genomic organization of the human catechol O-methyltransferase gene and its expression from two distinct promoters. *Eur J Biochem*, **223**, 1049-1059.

7. Bai, H.W., Shim, J.Y., Yu, J. and Zhu, B.T. (2007) Biochemical and molecular modeling studies of the O-methylation of various endogenous and exogenous catechol substrates catalyzed by recombinant human soluble and membrane-bound catechol-O-methyltransferases. *Chem Res Toxicol*, **20**, 1409-1425.

8. Vidgren, J., Svensson, L.A. and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase. *Nature*, **368**, 354-358.

9. Rutherford, K., Le Trong, I., Stenkamp, R.E. and Parson, W.W. (2008) Crystal structures of human 108V and 108M catechol O-methyltransferase. *J. Mol. Biol.*, **380**, 120-130.

10. Roca, M., Marti, S., Andres, J., Moliner, V., Tunon, I., Bertran, J. and Williams, I.H. (2003) Theoretical modeling of enzyme catalytic power: analysis of "cratic" and electrostatic factors in catechol O-methyltransferase. *J. Am. Chem. Soc.*, **125**, 7726-7737.

11. Kuhn, B. and Kollman, P.A. (2000) QM-FE and molecular dynamics calculations on catechol O-methyltransferase: free energy of activation in the enzyme and in aqueous

solution and regioselectivity of the enzyme-catalyzed reaction. *J. Am. Chem. Soc.*, **122**, 2586-2596.

12. Lotta, T., Vidgren, J., Tilgmann, C., Ulmanen, I., Melen, K., Julkunen, I. and Taskinen, J. (1995) Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry*, **34**, 4202-4210.

13. Diatchenko, L., Slade, G.D., Nackley, A.G., Bhalang, K., Sigurdsson, A., Belfer, I., Goldman, D., Xu, K., Shabalina, S.A., Shagin, D. *et al.* (2005) Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum Mol Genet*, **14**, 135-143.

14. Zhu, B.T. (2002) Catechol-O-Methyltransferase (COMT)-mediated methylation metabolism of endogenous bioactive catechols and modulation by endobiotics and xenobiotics: importance in pathophysiology and pathogenesis. *Curr Drug Metab*, **3**, 321-349.

15. Nackley, A.G., Shabalina, S.A., Tchivileva, I.E., Satterfield, K., Korchynskyi, O., Makarov, S.S., Maixner, W. and Diatchenko, L. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*, **314**, 1930-1933.

16. Rutherford, K., Alphandery, E., McMillan, A., Daggett, V. and Parson, W.W. (2008) The V108M mutation decreases the structural stability of catechol O-methyltransferase. *Biochim Biophys Acta*, **1784**, 1098-1105.

17. Nackley, A.G., Tan, K.S., Fecho, K., Flood, P., Diatchenko, L. and Maixner, W. (2007) Catechol-O-methyltransferase inhibition increases pain sensitivity through activation of both beta2- and beta3-adrenergic receptors. *Pain*, **128**, 199-208.

18. Vargas-Alarcon, G., Fragoso, J.M., Cruz-Robles, D., Vargas, A., Lao-Villadoniga, J.I., Garcia-Fructuoso, F., Ramos-Kuri, M., Hernandez, F., Springall, R., Bojalil, R. *et al.* (2007) Catechol-O-methyltransferase gene haplotypes in Mexican and Spanish patients with fibromyalgia. *Arthritis Res Ther*, **9**, R110.

19. Lee, P.J., Delaney, P., Keogh, J., Sleeman, D. and Shorten, G.D. (2010) Catecholamine-o-methyltransferase polymorphisms are associated with postoperative pain intensity. *Clin J Pain*, **27**, 93-101.

20. Bialecka, M., Kurzawski, M., Klodowska-Duda, G., Opala, G., Tan, E.K. and Drozdzik, M. (2008) The association of functional catechol-O-methyltransferase haplotypes with risk of Parkinson's disease, levodopa treatment response, and complications. *Pharmacogenet Genomics*, **18**, 815-821.

21. Liu, B., Li, J., Yu, C., Li, Y., Liu, Y., Song, M., Fan, M., Li, K. and Jiang, T. (2010) Haplotypes of catechol-O-methyltransferase modulate intelligence-related brain white matter integrity. *Neuroimage*, **50**, 243-249.

22. Tsao, D., Shabalina, S.A., Gauthier, J., Dokholyan, N.V. and Diatchenko, L. (2011) Disruptive mRNA folding increases translational efficiency of catechol-O-methyltransferase variant. *Nucleic Acids Res*.

23. Tsao, D., Liu, S. and Dokholyan, N.V. (2011) Regioselectivity of catechol O-methyltransferase confers enhancement of catalytic activity. *Chem Phys Lett*, **506**, 135-138.

24. Barnett, J.H., Heron, J., Goldman, D., Jones, P.B. and Xu, K. (2009) Effects of catechol-O-methyltransferase on normal variation in the cognitive function of children. *Am J Psychiatry*, **166**, 909-916.

25. Ducci, F. and Goldman, D. (2008) Genetic approaches to addiction: genes and alcohol. *Addiction*, **103**, 1414-1428.

26. Yin, S., Biedermannova, L., Vondrasek, J. and Dokholyan, N.V. (2008) MedusaScore: an accurate force field-based scoring function for virtual drug screening. *J Chem Inf Model*, **48**, 1656-1662.

27. Ding, F., Yin, S. and Dokholyan, N.V. (2010) Rapid flexible docking using a stochastic rotamer library of ligands. *J Chem Inf Model*, **50**, 1623-1632.

28. Lee, W.J., Shim, J.Y. and Zhu, B.T. (2005) Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids. *Mol Pharmacol*, **68**, 1018-1030.

29. Masri, M.S., Robbins, D.J., Emerson, O.H. and Deeds, F. (1964) Selective Para- or Meta-O-Methylation with Catechol O-Methyl Transferase from Rat Liver. *Nature*, **202**, 878-879.

30. Zheng, Y. and Bruice, T.C. (1997) A theoretical examination of the factors controlling the catalytic efficiency of a transmethylation enzyme: catechol O-methyltransferase. *J. Am. Chem. Soc.*, **119**, 8137-8145.

31. Papoian, G.A., DeGrado, W.F. and Klein, M.L. (2003) Probing the configurational space of a metalloprotein core: an ab initio molecular dynamics study of Duo Ferro 1 binuclear Zn cofactor. *J. Am. Chem. Soc.*, **125**, 560-569.

32. Case, D.A., Darden, T.A., Cheatham, T.E., III,, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M. *et al.* (2006) AMBER 9. *AMBER 9*.

33. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J., J. A.;, Vreven, T., Kudin, K.N., Burant, J.C. *et al.* (2004) Gaussian 03, Revision C.02. *Gaussian 03, Revision C.02*.

34. Liu, S. (2009) Conceptual density functional theory and some recent developments. *Acta Phys. Chim. Sin.*, **25**.

35. Liu, S., Hu, H. and Pedersen, L.G. (2010) Steric, quantum, and electrostatic effects on S(N)2 reaction barriers in gas phase. *J. Phys. Chem. A*, **114**, 5913-5918.

36. Lachman, H.M. (2008) Does COMT val158met affect behavioral phenotypes: yes, no, maybe? *Neuropsychopharmacology*, **33**, 3027-3029.

37. Voelker, P., Sheese, B.E., Rothbart, M.K. and Posner, M.I. (2009) Variations in catechol-O-methyltransferase gene interact with parenting to influence attention in early development. *Neuroscience*, **164**, 121-130.

38. George, S.Z., Dover, G.C., Wallace, M.R., Sack, B.K., Herbstman, D.M., Aydog, E. and Fillingim, R.B. (2008) Biopsychosocial influence on exercise-induced delayed onset muscle soreness at the shoulder: pain catastrophizing and catechol-o-methyltransferase (COMT) diplotype predict pain ratings. *Clin J Pain*, **24**, 793-801.

39. George, S.Z., Wallace, M.R., Wright, T.W., Moser, M.W., Greenfield, W.H., 3rd, Sack, B.K., Herbstman, D.M. and Fillingim, R.B. (2008) Evidence for a biopsychosocial influence on shoulder pain: pain catastrophizing and catechol-O-methyltransferase (COMT) diplotype predict clinical pain ratings. *Pain*, **136**, 53-61.

40. Ray, B.K., Lawson, T.G., Kramer, J.C., Cladaras, M.H., Grifo, J.A., Abramson, R.D., Merrick, W.C. and Thach, R.E. (1985) ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem*, **260**, 7651-7658.

41. Kozak, M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol*, **9**, 5134-5142.

42. Somogyi, P., Jenner, A.J., Brierley, I. and Inglis, S.C. (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol*, **13**, 6931-6940.

43. Hansen, T.M., Reihani, S.N., Oddershede, L.B. and Sorensen, M.A. (2007) Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc Natl Acad Sci U S A*, **104**, 5830-5835.

44. Mahen, E.M., Watson, P.Y., Cottrell, J.W. and Fedor, M.J. (2010) mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol*, **8**, e1000307.

45. Takyar, S., Hickerson, R.P. and Noller, H.F. (2005) mRNA helicase activity of the ribosome. *Cell*, **120**, 49-58.

46. Jackson, R.J., Hellen, C.U. and Pestova, T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, **11**, 113-127.

47. Shabalina, S.A., Ogurtsov, A.Y. and Spiridonov, N.A. (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res*, **34**, 2428-2437.

48. Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y. and Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103-107.

49. Shabalina, S.A., Ogurtsov, A.Y., Rogozin, I.B., Koonin, E.V. and Lipman, D.J. (2004) Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Res*, **32**, 1774-1782.

50. Chen, H., Pomeroy-Cloney, L., Bjerknes, M., Tam, J. and Jay, E. (1994) The influence of adenine-rich motifs in the 3' portion of the ribosome binding site on human IFN-gamma gene expression in Escherichia coli. *J Mol Biol*, **240**, 20-27.

51. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13-37.

52. Kozak, M. (1979) Inability of circular mRNA to attach to eukaryotic ribosomes. *Nature*, **280**, 82-85.

53. Griswold, K.E., Mahmood, N.A., Iverson, B.L. and Georgiou, G. (2003) Effects of codon usage versus putative 5'-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm. *Protein Expr Purif*, **27**, 134-142.

54. Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255-258.

55. Gu, W., Zhou, T. and Wilke, C.O. (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol*, **6**, e1000664.

56. Ding, F., Sharma, S., Chalasani, P., Demidov, V.V., Broude, N.E. and Dokholyan, N.V. (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA*, **14**, 1164-1173.

57. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

58. Ogurtsov, A.Y., Shabalina, S.A., Kondrashov, A.S. and Roytberg, M.A. (2006) Analysis of internal loops within the RNA secondary structure in almost quadratic time. *Bioinformatics*, **22**, 1317-1324.

59. Mathews, D.H., Burkard, M.E., Freier, S.M., Wyatt, J.R. and Turner, D.H. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, **5**, 1458-1469.

60. Kondrashov, A.S. and Shabalina, S.A. (2002) Classification of common conserved sequences in mammalian intergenic regions. *Hum Mol Genet*, **11**, 669-674.

61. Andersen, H.C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, **72**, 2384-2393.

62. Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E. and Shakhnovich, E.I. (1998) Discrete molecular dynamics studies of the folding of a protein-like model. *Fold Des*, **3**, 577-587.

63. Zhou, Y. and Karplus, M. (1997) Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.*, **107**, 10691-10708.

64. Sugita, Y. and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **314**, 141-151.

65. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A. and Rosenberg, J.M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, **13**, 1011-1021.

66. de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. and Vogel, C. (2009) Global signatures of protein and mRNA expression levels. *Mol Biosyst*, **5**, 1512-1526.

67. Seo, S.W., Yang, J. and Jung, G.Y. (2009) Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. *Biotechnol Bioeng*, **104**, 611-616.

68. Zhang, F., Saha, S., Shabalina, S.A. and Kashina, A. (2010) Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science*, **329**, 1534-1537.

69. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.

70. Matveeva, O., Nechipurenko, Y., Rossi, L., Moore, B., Saetrom, P., Ogurtsov, A.Y., Atkins, J.F. and Shabalina, S.A. (2007) Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res*, **35**, e63.

71. Morgan, S.R. and Higgs, P.G. (1998) Barrier heights between ground states in a model of RNA secondary structure. *J. Phys. A: Math. Gen.*, **31**, 3153-3170.

72. Hyeon, C. and Thirumalai, D. (2003) Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc Natl Acad Sci U S A*, **100**, 10249-10253.

73. Solomatin, S.V., Greenfeld, M., Chu, S. and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681-684.

74. Tuller, T., Waldman, Y.Y., Kupiec, M. and Ruppin, E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A*, **107**, 3645-3650.

75. Gonzalez de Valdivia, E.I. and Isaksson, L.A. (2004) A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in Escherichia coli. *Nucleic Acids Res*, **32**, 5198-5205.

76. Waldman, Y.Y., Tuller, T., Sharan, R. and Ruppin, E. (2009) TP53 cancerous mutations exhibit selection for translation efficiency. *Cancer Res*, **69**, 8807-8813.

77. Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I. and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, **141**, 344-354.

78. Dittmar, K.A., Goodenbour, J.M. and Pan, T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet*, **2**, e221.

79. Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489-499.

80. Darnell, J.C., Fraser, C.E., Mostovetsky, O., Stefani, G., Jones, T.A., Eddy, S.R. and Darnell, R.B. (2005) Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev*, **19**, 903-918.

81. Li, Y., Lin, L. and Jin, P. (2008) The microRNA pathway and fragile X mental retardation protein. *Biochim Biophys Acta*, **1779**, 702-705.

82. Plante, I., Davidovic, L., Ouellet, D.L., Gobeil, L.A., Tremblay, S., Khandjian, E.W. and Provost, P. (2006) Dicer-Derived MicroRNAs Are Utilized by the Fragile X Mental Retardation Protein for Assembly on Target RNAs. *J Biomed Biotechnol*, **2006**, 64347.

83. Schul, R. and Frenk, H. (1991) The role of serotonin in analgesia elicited by morphine in the periaqueductal gray matter (PAG). *Brain Res*, **553**, 353-357.

84. Roberts, M.H. (1984) 5-Hydroxytryptamine and antinociception. *Neuropharmacology*, **23**, 1529-1536.

85. Davies, J.E., Marsden, C.A. and Roberts, M.H. (1983) Hyperalgesia and the reduction of monoamines resulting from lesions of the dorsolateral funiculus. *Brain Res*, **261**, 59-68.

86. Smith, A.J. (1998) The analgesic effects of selective serotonin reuptake inhibitors. *J Psychopharmacol*, **12**, 407-413.

87. Coluzzi, F. and Mattia, C. (2005) Mechanism-based treatment in chronic neuropathic pain: the role of antidepressants. *Curr Pharm Des*, **11**, 2945-2960.

88. Banks, W.A., Kastin, A.J., Trentman, T.L., Haynes, H.S., Johnson, B.G. and Galina, Z.H. (1988) Mediation of serotonin-induced analgesia by the 5HT2 receptor in the pentobarbital anesthetized mouse model. *Brain Res Bull*, **21**, 887-891.

89. Taiwo, Y.O. and Levine, J.D. (1992) Serotonin is a directly-acting hyperalgesic agent in the rat. *Neuroscience*, **48**, 485-490.

90. Segall, S.K., Nackley, A.G., Diatchenko, L., Lariviere, W.R., Lu, X., Marron, J.S., Grabowski-Boase, L., Walker, J.R., Slade, G., Gauthier, J. *et al.* (2010) Comt1 genotype and expression predicts anxiety and nociceptive sensitivity in inbred strains of mice. *Genes Brain Behav*, **9**, 933-946.

91. Khare, S.D., Caplow, M. and Dokholyan, N.V. (2004) The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A*, **101**, 15094-15099.

92. Chen, J., Lipska, B.K., Halim, N., Ma, Q.D., Matsumoto, M., Melhem, S., Kolachana, B.S., Hyde, T.M., Herman, M.M., Apud, J. *et al.* (2004) Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am J Hum Genet*, **75**, 807-821.

93. Lin, S.Y., Chang, W.J., Lin, C.S., Huang, C.Y., Wang, H.F. and Sun, W.H. (2011) Serotonin receptor 5-HT2B mediates serotonin-induced mechanical hyperalgesia. *J Neurosci*, **31**, 1410-1418.

94. Messing, R.B. and Lytle, L.D. (1977) Serotonin-containing neurons: their possible role in pain and analgesia. *Pain*, **4**, 1-21.