General and Comparative Endocrinology 215 (2015) 106–116

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

Comparative Molecular Analyses

Structural libraries of protein models for multiple species to understand evolution of the renin-angiotensin system

Jeremy W. Prokop [a,b,*], Victoria Petri [a], Mary E. Shimoyama [a,c], Ingrid K.M. Watanabe [d], Dulce E. Casarini [d], Thomas C. Leeper [e], Stephanie M. Bilinovich [e], Howard J. Jacob [a,b,f], Robson A.S. Santos [g], Almir S. Martins [g], Fabiano C. Araujo [h], Fernando M. Reis [h], Amy Milsted [i]

[a] Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI, USA
[b] Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA
[c] Department of Surgery, Medical College of Wisconsin, Milwaukee, WI, USA
[d] Nephrology Division, Department of Medicine, Federal University of Sao Paulo, Sao Paulo, SP, Brazil
[e] Department of Chemistry, The University of Akron, Akron, OH, USA
[f] Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA
[g] Departamento de Fisiologia e Biofísica, Instituto de Ciências Biológicas, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[h] National Institute of Science and Technology in Molecular Medicine – Department of Obstetrics and Gynecology, School of Medicine, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[i] Department of Biology, The University of Akron, Akron, OH, USA

ARTICLE INFO

Article history:
Available online 26 September 2014

Keywords:
Sequence-to-structure-to-function analysis
Rat genetics
Comparative modeling
Renin-angiotensin system
Angiotensin peptides

ABSTRACT

The details of protein pathways at a structural level provides a bridge between genetics/molecular biology and physiology. The renin-angiotensin system is involved in many physiological pathways with informative structural details in multiple components. Few studies have been performed assessing structural knowledge across the system. This assessment allows use of bioinformatics tools to fill in missing structural voids. In this paper we detail known structures of the renin-angiotensin system and use computational approaches to estimate and model components that do not have their protein structures defined. With the subsequent large library of protein structures, we then created a species specific protein library for human, mouse, rat, bovine, zebrafish, and chicken for the system. The rat structural system allowed for rapid screening of genetic variants from 51 commonly used rat strains, identifying amino acid variants in angiotensinogen, ACE2, and AT1b that are in contact positions with other macromolecules. We believe the structural map will be of value for other researchers to understand their experimental data in the context of an environment for multiple proteins, providing pdb files of proteins for the renin-angiotensin system in six species. With detailed structural descriptions of each protein, it is easier to assess a species for use in translating human diseases with animal models. Additionally, as whole genome sequencing continues to decrease in cost, tools such as molecular modeling will gain use as an initial step in designing efficient hypothesis driven research, addressing potential functional outcomes of genetic variants with precompiled protein libraries aiding in rapid characterizations.

1. Introduction

Whole genome sequencing is in an age of rapid expansion, with the estimate of wet lab costs to dip below $1000 in the very near future through the release of technology such as the HiSeq X machine. This will require new tools to help in understanding how genetic variants are connected to changes in physiology and how to use the increasing number of sequenced species to better understand the mechanisms of evolution. Molecular modeling and dynamics will provide some of the necessary additions to

Abbreviations: RAS, renin-angiotensin system; AGT, angiotensinogen; GPCRs, G-protein coupled receptors; SSF, sequence-to-structure-to-function; Ang, angiotensin; PRR, (pro)renin receptor; pdb, Protein Data Bank; ACE, angiotensin converting enzyme; md, molecular dynamic; PRCP, Lysosomal Pro-X carboxypeptidase; GST, glutathione S-transferase; RMSD, root-mean squared deviation.

* Corresponding author at: Human Molecular and Genetics Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA. Fax: +1 414 456 6516.

E-mail address: jprokop54@gmail.com (J.W. Prokop).

http://dx.doi.org/10.1016/j.ygcen.2014.09.010
0016-6480/

the toolkit for interpretation of genome variation. Currently whole exome sequencing, an analysis of the protein coding genes, is the least expensive toolset for understanding human genetic variants. Combining the sequencing of protein coding genes with evolutionary analysis of diverse species and molecular modeling tools can move simple identification into functional prediction. More importantly, with the proper tools available, this approach can be done with speed, and improve our ability to address variants in entire protein pathways. Having structural protein libraries of pathways already assembled and publically available is one way to increase the speed in identifying functional outcomes of genetic diversity. Utilizing such an approach, we analyzed the renin-angiotensin system (RAS), building protein libraries for human, mouse, rat, bovine, zebrafish, and chicken. The structural library was then used to analyze 51 rat genomes of strains commonly used as animal models for the genetics of cardiovascular disease.

The RAS is a complex pathway that has important roles in many cardiovascular, renal, and endocrine processes and cell proliferation. The pathway consists of the protein angiotensinogen (AGT) that is cleaved by various enzymes, generating peptide fragments (angiotensin, Ang) that bind and activate G-protein coupled receptors (Fig. 1). The structures of many of the components in the RAS are known, with several protein–protein or inhibitor bound structures reported (Table S1). Efforts at inhibiting multiple steps of the RAS pathway have been successful in treating hypertension in millions of patients. However, our current knowledge of this pathway at the molecular and DNA sequence variant level is still far from complete and thus many potential mechanisms of the RAS may not be defined. In this paper we integrate all known 3D protein structures for components of the RAS with proteins (and peptides) that do not have a reported structure. For this latter case we use *in silico* analyses (molecular modeling, AutoDock prediction, and molecular dynamic simulations) to help elucidate a full structural analysis of the RAS (Fig. 1). We hypothesize that understanding the complete sequence-to-structure-to-function (SSF) for the RAS will serve as a tool in identifying genetic variants that may alter protein function while also offering a toolset in studying animal models for human diseases.

The activation of the RAS begins with the expression of AGT, which can exist in either a reduced or oxidized state, with different production rates of Ang peptides from the two forms (Zhou et al., 2010). The oxidized AGT is rapidly processed by the enzyme renin to produce a ten amino acid fragment known as angiotensin I (Ang I). Renin is first translated as an inactive protein (zymogen) with a propeptide (Sealey et al., 1980). This propeptide disrupts beta sheet packing within renin causing a steric block of the active site (Morales et al., 2012) resulting in low levels of catalytic activity in cleaving AGT (Heinrikson et al., 1989). Activation of the zymogen's propeptide through cleavage by various enzymes (Reudelhuber et al., 1994), low pH (Danser and Deinum, 2005), cold temperatures (Danser and Deinum, 2005), or binding of prorenin to the (pro)renin receptor (PRR) can all increase the catalytic activity of renin on AGT, resulting in the release of Ang I. Analysis of the concentrations of renin vs. prorenin circulating in the blood has suggested upwards of a 10-fold higher level of prorenin, implicating potential tissue specific roles of prorenin activation by PRR (Zhuo, 2011). In addition, prorenin binding to PRR activates intracellular pathways such as ERK 1/2 with proposed roles in cranial RAS signaling (Nguyen, 2011; Nguyen et al., 2002). The process of activation of renin and the impact of oxidized and reduced AGT can be seen at https://www.youtube.com/watch?v=RsEEs5WFkSQ&feature=youtu.be.

Once Ang I is formed, it is cleaved by various enzymes to produce peptides of additional sizes. These enzymes include ACE (PDB files 2c6f and 1o8a), ACE2 (1r42), Neprilysin (NEP, 1r1h), and the Lysosomal Pro-X carboxypeptidase (PRCP, 3n2z). ACE contains two catalytic active sites (Soubrier et al., 1988) highly documented for the production of the Ang II peptide. The two domains show different Cl^- ion concentration in activation (Wei et al., 1991) with an additive effect when both domains are present (Marcic et al., 2000). Ang II can further be processed in two directions. Aminopeptidase A converts Ang II into Ang III by cleaving off the first amino acid (Asp1), which is further processed by Aminopeptidase N into Ang IV with removal of Arg2. A homologous enzyme to ACE, known as ACE2, efficiently cleaves Ang II to make Ang-(1–7) (Rice et al., 2004). Ang-(1–7) can be converted into the peptide Alamandine by cleavage of the amino acid 1 side chain resulting in an alanine at this site (Lautner et al., 2013).

The production of these peptides contributes to the two arms of the RAS pathway, that of Ang II eliciting vasoconstriction or that of Ang-(1–7) functioning in opposition. Numerous studies have addressed the balance between these two arms of the pathway in diseases from hypertension (Brosnihan et al., 2005) to cancer (Ager et al., 2008). Ang II activates the AT1 receptor in a two-step process eliciting increase in vasoconstriction, angiogenic, proliferative effects (Hines et al., 2003; Holloway et al., 2002; Hunyady et al., 2003; Kobilka and Deupi, 2007; Noda et al., 1996; Prokop et al., 2013; Vauquelin and Van Liefde, 2005). Ang III preferentially

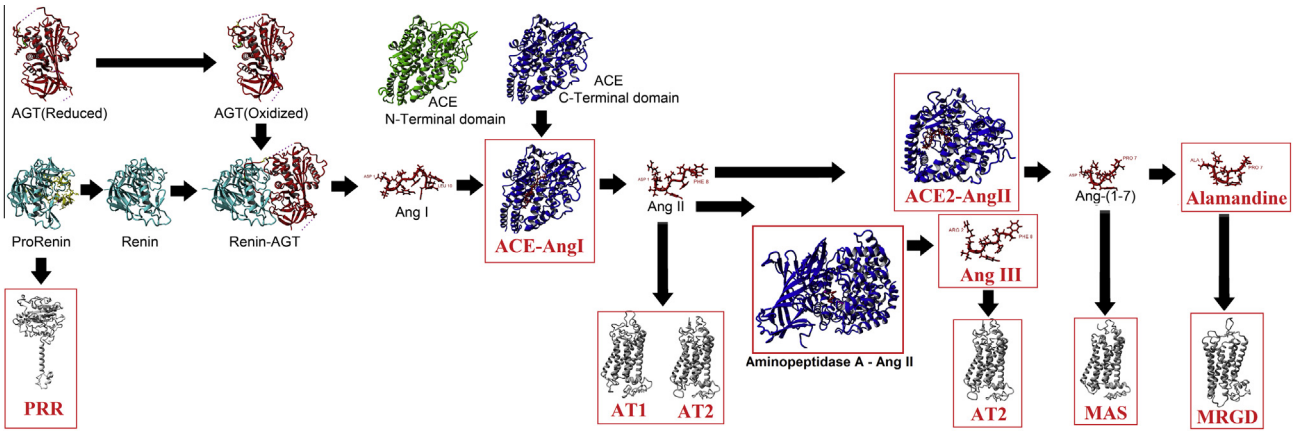

Fig. 1. Multiple components of the renin-angiotensin system. Structural analysis of the RAS with components that are characterized using the *in silico* approaches in this paper boxed in red. Biochemically determined structures are known for AGT (reduced/oxidized), Prorenin, Renin, Renin-AGT, Ang I, Ang II, Ang-(1–7), ACE (N/C-terminal domains), and ACE 2. Models were used for the PRR, AT1, AT2, MAS, MRGD, ACE-Ang I, Aminopeptidase A-Ang II, and ACE2–Ang II. In addition to the models shown here we have generated potential structures for Ang-Neprilysin and Ang-PRCP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)