

Stony Brook University



OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Structural and energetic determinants of function in the heterotrimeric G-proteins

A Dissertation Presented

by

Noel Carrascal

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics & Statistics

(Computational Biology)

Stony Brook University

December 2011

Stony Brook University

The Graduate School

Noel Carrascal

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

David Francis Green – Dissertation Advisor
Associate Professor Applied Mathematics and Statistics

Yuefan Deng – Chairperson of Defense
Professor Applied Mathematics and Statistics

Robert C. Rizzo – Dissertation Co-adviser
Associate Professor Applied Mathematics and Statistics

Suzanne Scarlata
Professor Physiology & Biophysics

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Structural and energetic determinants of function in the Heterotrimeric G-Proteins

by

Noel Carrascal

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Computational Biology)

Stony Brook University

2011

The heterotrimeric G-proteins are important in signal transduction that triggers important biological functions in cells of living organisms. The use of theoretical models to study these proteins gives additional insight into mechanisms of signal transduction and molecular recognition events in the cell. With that objective in mind, molecular dynamics simulations of a heterotrimeric G-protein reveal structural and energetic differences among the states of its signaling cycle. The simulations are stripped explicit waters and analyzed in search of determinants of function with the help of backbone clustering and continuum electrostatic models such as Poisson–Boltzmann (PB) and generalized-Born (GB). The energies are broken down into contributions from individual components corresponding to the carbonyl and amino part of the backbone as well as the side chain of each residue. The results allow the identification of components that are important for folding and binding. Two interdependent results emerge from this research: continuum electrostatic models facilitate the understanding of biological systems from a theoretical point of view; and the heterotrimeric G-protein in its three states serves as a biologically important model for evaluating electrostatic continuum models. As a preliminary work, an evaluation of the GB continuum model was done by measuring entropic differences in free energy of binding from simulations in continuum and explicit solvent models; this supports the use of explicit solvent for molecular dynamics and GB for measuring

electrostatic binding free energies. Explicit water for simulation and continuum electrostatic free energy calculations—with GB and PB—were further tested by simulating a heterotrimeric G-protein in three different states for 100ns of simulation time. The degree to which GB approximates PB was quantified for total binding energies as well as for individual components' contributions to binding energies. GB is corroborated as a reasonably good approximation to PB, but it appears to be most useful as a filter of relevant components to be calculated with the more reliable PB method. The simulations of the heterotrimeric G-proteins were extended to 608ns, and further analysis included clustering and the calculations of folding free energies in addition to binding free energies. Cluster analysis was performed on the backbone of the α -subunit revealing structural differences in sub domains and loop regions of the heterotrimeric G-protein in its different states. Differences in binding energies from one state of the heterotrimeric G-protein to the other were used to identify and structurally characterize the interaction of important components. At the end of this work, a detailed description of structural and energetic determinants of function in the heterotrimeric G-protein is given at the atomic level. This information gives important insight into the changes that the heterotrimeric G-protein undergo during signal transduction from theoretical models.

Table of Contents

1. Introduction.....	1
1.1 Heterotrimeric G-protein function.....	1
1.1.1 General function.....	1
1.1.2 Structural features.....	3
1.2 Computational methods.....	5
1.2.1 Molecular dynamics simulations.....	5
1.2.2 Computational studies of G-proteins.....	7
1.3 Structural and energetic analysis from trajectories.....	8
1.3.1 Comparing ensembles generated with implicit and explicit water models help in the interpretation of free energies of binding.....	8
1.3.2 Energetic decomposition with the generalized-Born and Poisson–Boltzmann solvent models: lessons from association of G-protein components.....	9
1.3.3 Structural motives of heterotrimeric G-protein subunits and domains from backbone clustering.....	9
1.3.4 Energetic decomposition of binding and stability of selected components from extended molecular dynamic simulations.....	10
2. Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13.....	11
2.1 Abstract.....	11
2.2 Introduction.....	11
2.3 Theory.....	12
2.4 Methods.....	13
2.5 Results.....	15
2.6 Conclusion.....	18
3. Energetic decomposition with the generalized-Born and Poisson–Boltzmann solvent models: lessons from association of G-protein components.....	20
3.1 Abstract.....	20
3.2 Introduction.....	21
3.3 Theory.....	23

3.3.1	The linearized-Poisson–Boltzmann model is pairwise decomposable.....	23
3.3.2	Free-energy differences in the linearized-Poisson–Boltzmann model.....	26
3.3.3	Energetic decomposition in the linearized-Poisson–Boltzmann model.....	27
3.3.4	Energetic decomposition in the generalized-Born model.....	30
3.4	Methods.....	32
3.4.1	Structure preparation.....	32
3.4.2	Molecular dynamics.....	32
3.4.3	Poisson–Boltzmann calculations.....	32
3.4.4	Generalized-Born calculations.....	33
3.4.5	Component analysis.....	33
3.5	Results.....	34
3.5.1	Electrostatic binding free energies computed with generalized Born and with linearized Poisson–Boltzmann are well-correlated.....	35
3.5.2	Generalized-Born overestimates net electrostatic binding free energies.....	35
3.5.3	Generalized-Born underestimates differences in net electrostatic binding free energies between configurations.....	36
3.5.4	Energetic components show highly variable correlations between GB and LPB models.....	37
3.5.5	Direct, solvent-screened Coulombic interactions show bimodal correlations.....	38
3.5.6	Inconsistencies between GB and PB are group-specific.....	40
3.5.7	Contributions of a buried salt bridge are missed by GB.....	42
3.5.8	Deviations between PB and GB are larger for nucleotide binding....	42
3.5.9	Moderate-range, action-at-a-distance interactions are poorly captured by GB.....	48
3.5.10	Structurally similar interactions are reproduced between systems with both methods.....	48

3.5.11	Quantification of energetic differences between systems can be very model dependent.....	50
3.5.12	Generalized-Born is an effective filter for the selection of important groups.....	53
3.5.13	Salt effects are well correlated between LPB and GB models, when significant.....	54
3.5.14	Alternative approaches and directions of study.....	56
3.6	Conclusion.....	58
4.	Structural analysis of the heterotrimeric G-protein in three different states from long molecular dynamic simulations.....	60
4.1	Abstract.....	60
4.2	Introduction.....	61
4.3	Methods.....	64
4.3.1	Structure preparation.....	64
4.3.2	Molecular dynamics.....	64
4.3.3	Preparation and alignment of trajectories for comparison.....	65
4.3.4	Two dimensional root mean square deviation (2D-RMSD).....	65
4.3.5	Dendrograms from 2D-RMSDs.....	66
4.3.6	Average and maximum cluster RMSDs.....	67
4.3.7	Clustering of regions from separate simulations.....	67
4.4	Results and discussion.....	68
4.4.1	Motions between sub-domains.....	68
4.4.1.1	N-terminus domain.....	68
4.4.1.2	GTPase domain.....	69
4.4.1.3	The α -helical domain.....	70
4.4.2	Motions in switch regions.....	72
4.4.2.1	Switch region one.....	73
4.4.2.2	Switch region two.....	75
4.4.2.3	Switch region three.....	78
4.4.2.4	Switch region four.....	80

4.4.3	Comparison of B-factors between crystallographic structures and simulations.....	82
4.5	Conclusion.....	84
5.	Energetic analysis of relevant components from extended molecular dynamic simulations of heterotrimeric G-proteins.....	86
5.1	Abstract.....	86
5.2	Introduction.....	86
5.3	Methods.....	90
5.3.1	Poisson–Boltzmann calculations.....	90
5.3.2	Generalized-Born calculations.....	90
5.3.3	Tabulation and ranking of components within the cutoff.....	91
5.4	Results and discussion.....	91
5.4.1	Important components for binding.....	91
5.4.1.1	Nucleotide-protein binding energy.....	91
5.4.1.1.1	Nucleotide-protein binding group one.....	93
5.4.1.1.2	Nucleotide-protein binding group two.....	94
5.4.1.1.3	Nucleotide-protein binding group three.....	96
5.4.1.1.4	Nucleotide-protein binding group four.....	98
5.4.1.1.5	Nucleotide-protein binding group five.....	100
5.4.1.2	Binding energies between α and β	102
5.4.1.2.1	Protein–protein group one.....	103
5.4.1.2.2	Protein–protein group two.....	104
5.4.1.2.3	Protein–protein group three.....	105
5.4.1.2.4	Protein–protein group four.....	105
5.4.1.2.5	Protein–protein group five.....	106
5.4.1.2.6	Protein–protein group six.....	106
5.4.1.2.7	Protein–protein group seven.....	107
5.4.2	Important components for folding.....	108
5.4.2.1	Folding group one.....	109
5.4.2.2	Folding group two.....	110
5.4.2.3	Folding group three.....	111

5.4.2.4 Folding group four.....	114
5.5 Conclusion.....	117
6. Conclusions and future direction.....	118
6.1 Conclusion.....	118
6.2 Future directions.....	121
References.....	123
Appendix A: Supplementary material for chapter 3.....	134
Appendix B: Completing missing structural information in the crystallographic structures from homology modeling between the three systems.....	148

Acknowledgments

I would like to thank my adviser David F. Green for making me a better scientist, and making the work that I produce better with his corrections and suggestions. I also want to thank him for accepting me into his lab when it was very much in doubt that I could continue pursuing this doctorate degree.

To Robert C. Rizzo and Suzanne Scarlata for the opportunity to work in their labs, I want to thank them for such a great learning experience. Dr. Rizzo introduced me to the amazing world of computational drug designed by challenging me with projects for developing software for virtual drug screening. Dr. Scarlata gave me the support and opportunity to do experiments with all their intricacies and complexities.

I want to mention all of the lab mates with whom I shared office space for all these years. In no particular order: Sudipto Mukherjee, Ricky Chachra, Vadim Patsalo, Jonathan Cheng, Yukiji Fujimoto, Trent Balius, Loretta Au, Tao Jiang and Dustin Brody. It was with them that I exchanged ideas, knowledge and had a great friendship through the years.

Gratitude is expressed to the New York State Office of Science Technology and Academic Research, the Office of the Vice President of Research at Stony Brook, and the Computational Science Center at Brookhaven National Laboratory for support in the material produced in chapter 2.

I would like to thank Bruce Tidor for making the Integrated Continuum Electrostatic (ICE) software package available. D.F.G. also thanks the State University of New York, the College of Engineering and Applied Sciences, and the Department of Applied Mathematics and Statistics for financial support. I was partially supported by a Bridges to the Doctorate fellowship from the NSF-SUNY Alliance for Graduate Education and the Professoriate.

I am very grateful to the staff of the Applied Mathematics and Statistics department and Math tower for always being there when I needed help: Victor Poon, Christine Rota, Janice Hackney, Laurie Delassio, Laura Colucci and Lucrecia Arasena.

I want to thank my brothers for supporting me in this long expedition into the United States that is still going. I would not be here if it wasn't for my mother who put me at the

doorstep of this country, and for my father, who passed away, for giving me the dreams that keep me looking for a special destiny in my career and in life.

I would like to mention Stacey Massignan for her support during this last year of my PhD; my feelings for her were a source of motivation and enthusiasm during the long and difficult hours of writing this thesis.

CHAPTER 1

Introduction

1.1 Heterotrimeric G-protein function

1.1.1 General function

Guanine nucleotide binding proteins, such as the heterotrimeric G-proteins, are important in the transmission of extracellular signals into the cell. These proteins are commonly localized in the proximity of trans-membrane G-Protein Coupled Receptors (GPCRs), with post-translationally modified termini that allow them to remain attached to the cell membrane.(1) Heterotrimeric G-proteins and GPCRs interact in many different ways depending on sequence variability. At least 865 genes have been identified to code for GPCRs in the human genome; these receptors work together with heterotrimeric G-proteins to modulate the transduction of signals into the cell.(2) In addition to the large number of genes that code for GPCRs, there are also multiple genes that code for heterotrimeric G-proteins. There are at least 21 G_{α} , 6 G_{β} and 12 G_{γ} subunits of the heterotrimeric G-proteins in humans.(3) G_{α} is primarily used to classify heterotrimeric G-proteins based on sequence similarity. There are four distinct classes of G_{α} : G_{s} , G_{i} , G_{q} and G_{12} .(4) The large possible number of combinations between heterotrimeric G proteins and GPCRs allows for the modulation of a large variety of signal pathways, which activate the machinery and specific cellular function in a variety of organisms that range in complexity from unicellular yeast to cells in tissues of humans and other mammals.(5-6)

Heterotrimeric G-proteins wait for activating signals from GPCRs that are passed into the cell through a series of protein-protein interactions and conformational motions that, to some extent, have been characterized experimentally.(7-9) Experiments suggest that upon GPCR activation, the heterotrimeric G-protein undergoes conformational changes that precede nucleotide exchange. After GDP is replaced with GTP, it was originally believed that the heterotrimeric G-protein would dissociate into the G_{α} and $G_{\beta\gamma}$ subunits; however, in a proposed second mode of signal transduction the heterotrimeric G-protein subunit simply rearranges rather than fully dissociating.(10-12)

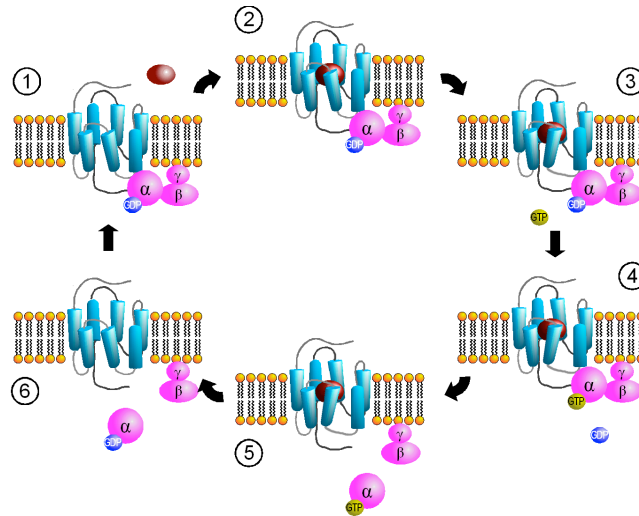


Figure 1.1. Example of a heterotrimeric G-protein signaling cycle. (1) GPCR and heterotrimeric G-protein wait for an activator. (2) a ligand binds to the GPCR. (3) Changes in the complex signals for nucleotide exchange. (4) GDP is replaced by GTP. (5) the α and $\beta\gamma$ dissociate. (6) α hydrolyzes GTP, turns it into GDP and the system goes back to state 1. (This figure was copied from Wikipedia under GNU-FDL licensing and was created by Sven Jähnichen)

In addition to the two possible modes of activation, dissociation or rearrangement, heterotrimeric G-proteins and GPCRs may work in a variety of combinations. The simplest mode of operation is a single heterotrimer with a single receptor. It is also possible to have GPCRs bound to their respective heterotrimers in a 1:1 stoichiometry in multiple pairs. Finally, a 2:1 stoichiometry has been experimentally observed that suggests that G_{α} and $G_{\beta\gamma}$ bind to different GPCRs, as opposed to a single heterotrimer per GPCR.⁽¹³⁾ In addition to the possible rearrangements between heterotrimers and GPCRs, there are many other proteins that interact in a concerted mode that aid signal transduction in ways depending on the surface of the cell environment and the tissue where the protein is present.

Some proteins involved in downstream signaling of G-proteins are G-protein Exchange Factors (GEFs), Regulators of G-protein signals (RGSs), and G-protein Interacting Proteins (GIPs). For heterotrimeric G-proteins, GPCRs correspond to the GEF that activates them. Downstream signaling after activation involves many types of GIPs that are involved in a large number of functions in different cells such as G-protein-gated ion channels found in neurons. Active heterotrimeric G-proteins regulate other proteins involved in processes such as metabolic enzymes, ion channels and transporters, transcriptional machinery, motility or contractility machinery and secretory machinery. Passing of the signal through the heterotrimeric G-protein

into the interior of the cells generates cascading signaling that bifurcates in patterns that generate complex reactions in the cell. Some of the outcomes of this downstream signaling are glucose metabolism, steroid production, pacemaker activity, chemotaxis, cardiac function, synaptic plasticity, and pituitary function. Therefore, heterotrimeric G-proteins are part of intricate large scale systems such as organism homeostasis, embryonic development, gonadal development and learning and memory.(4-5, 14-15) Despite of the diversity of functions, configurations and roles of the heterotrimeric G-proteins, they all have two important properties in common: they recognize GDP and GTP as well as their corresponding $G_{\beta\gamma}$.

1.1.2 Structural features

There are three crystal structures of the G-protein that are particularly interesting because they all have identical subunit sequence corresponding to $G_{i\alpha 1}$. Two of the crystal structures are monomers in the inactive (1GDD) and active (1GIA) states, and the other is the heterotrimer in the inactive state (1GP2).(16-18) The G_{α} subunits can be generally divided into three large domains as shown in Figure 1.2. Domain 1 interacts with G_{β} in the heterotrimeric state, and its N-terminus is modified to go into the cell membrane. Domain 2, also called the GTPase domain, is well conserved in all members of the heterotrimeric G-protein super family.(3) Domain 3 is an α helical bundle unique to G_{α} proteins that buries the nucleotide binding pocket into the core of the protein.(19)

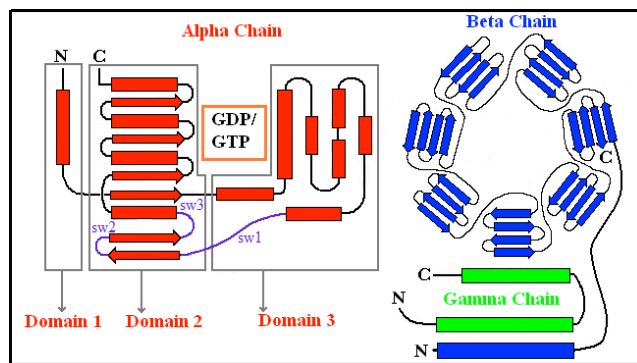


Figure 1.2. Topology of the heterotrimeric G-protein fold. Domains 1 and 3 are made of α helical structures and domain 2 is formed by a combination of α helical and β sheet structures. The location of the nucleotide is roughly between Domains 2 and 3. Switch regions 1, 2, 3 are shown in purple. The lengths of α helical, β sheets and loop regions are not drawn to scale. This schematization does not intend to show the binding orientation between α , β and γ subunits.

There are three switch regions that play a very important role in G_{α} specificity and affinity of binding to $G_{\beta\gamma}$. Conformational motions of these switch regions relative to each other,

and the GTPase domain, are believed to play an important role in modulating interactions with $G_{\beta\gamma}$.(20) Finally, G_{β} is a heptameric β -propeller with a helical domain at the N-terminus.

Figure 1.3 shows that the most notable differences between the G_i alpha 1 in the crystal structures is at the N-terminus; it bends towards the G_{α} domain in (1GDD), is missing in (1GIA), and is extended along $G_{\beta\gamma}$ in the heterotrimeric complex. The G_{γ} subunit interacts mostly with G_{β} by forming an inter-twined coil-coil structure at the helical region, and it is opposite in location to G_{α} relative to G_{β} . The crystal structure of the heterotrimer (1GP2) does not show any contact between the N-terminus of G_{α} and the C-terminus of G_{γ} ; however, their proximity indicates possible cooperation to penetrate into the membrane.(19)

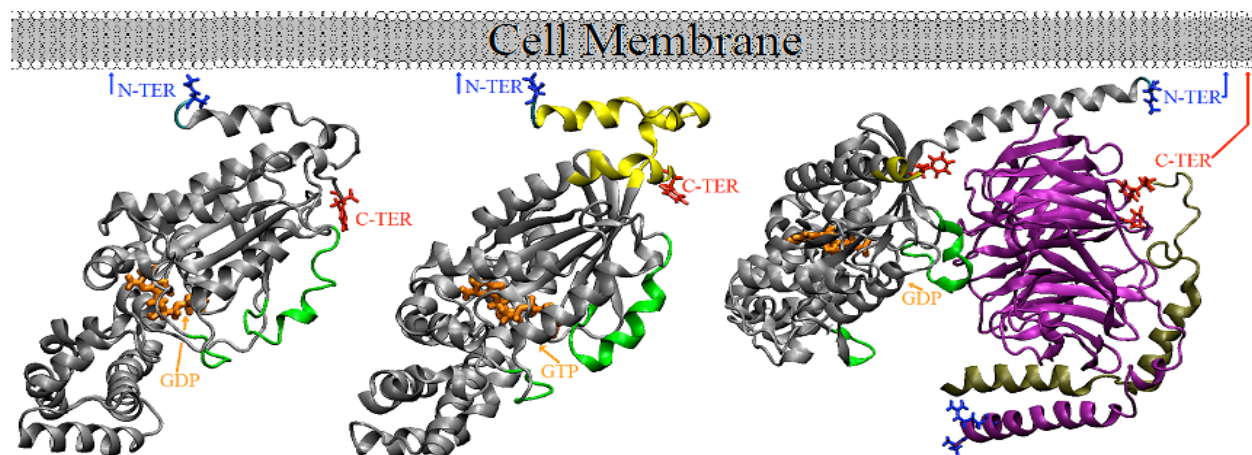


Figure 1.3. Structures of the three G-protein systems. From left to right: 1GDD, 1GIA and 1GP2. Added structural information is shown in yellow, green and cyan colors. The cell membrane is shown only for illustration purposes, and it is not included in the simulations.

A recently crystallized structure of the GPCR β_2 adrenergic receptor with a nucleotide-free G_s heterotrimer bound to it reveals new insights about the purported mechanism of function inferred from previous crystallographic structures and experimental observations. The most revealing observation in the heterotrimeric G-protein of this structure is a displacement of the α -helical domain relative to the GTPase domain. The involvement of these two domains in nucleotide binding, in addition to the range of motion of the α -helical domain, is a *smoking gun* that helps revealing the mechanism of nucleotide exchange that is fundamental in the heterotrimeric G-protein cycle.(21) The results of this work supports the importance in

characterizing the motions between the α -helical and GTPase domain in the structures of $G_{i\alpha 1}$ simulated here.

1.2 Computational methods

1.2.1 Molecular dynamic simulations

The relevance of molecular simulations to understanding protein function is now enhanced with the help of powerful supercomputers which allow for the study of larger proteins for longer periods of simulation time. The internal motions of atoms and protein domains as a function of time, and in their biological environment, cannot easily be observed with most experimental techniques. For this purpose, molecular dynamics simulations of large proteins have been done to reproduce and expand on experimental information.(22) Molecular dynamics simulations are based on strong theoretical foundations in classical physics that can be summarized in the form of the Hamiltonian:

$$\begin{aligned}
 V(r^n) = & \sum_{bond} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(nw - \gamma)) \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N (4\epsilon_{ij} [(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}})
 \end{aligned} \tag{1.1}$$

The Hamiltonian above is used to calculate the forces acting between atoms in the protein. The first two terms model harmonic oscillators with their corresponding constants k_i . In the first term in the equation, l_i corresponds to the distance between pairs of atoms and l_{i0} to a reference value. The second term is a summation over the difference between an angle (θ_i) and a reference angle (θ_{i0}) formed by consecutive atoms A-B-C, with A and C both bonded to B. The third term is a torsional potential of dihedral angles formed by four consecutive bonded atoms, and it is modeled with a sinusoidal function where n is the multiplicity (given by the number of minima as the bond is rotated through 360°), w is the dihedral angle and γ is the phase factor that determines where the dihedral angle passes through the minimum. The fourth term corresponds to non-bonded pair-wise interactions between all atoms in the protein. This term is formed by a Coulomb potential of electrostatic interactions, which is a function of distance between atoms r_{ij} , charge q and the permittivity of free space ϵ_0 , and a Lennard-Jones potential that accounts for van der Waals interactions, which is a function of depth of the potential well ϵ_{ij} , and σ_{ij} , is a

distance to establish the point at which the interparticle potential toggles from attractive to repulsive. The sum of these potentials in the equation gives the total energy of the system.

Physical models used in simulations have given structural biologists new insights about the dynamics of proteins.(22) The wide range of applications of these models include entropy calculations by Brooks and Karplus in 1983 and the estimation of structural populated states from long simulations by Shaw in 2010; both studies were done on the bovine pancreatic trypsin inhibitor.(23-25) The simulation of the trypsin inhibitor by Shaw was done for a single trajectory with a specially designed supercomputer that allows simulations to be extended to milliseconds of simulation time. Supercomputers are still of limited access to most scientists. For that reason, methods to assemble discrete molecular dynamic simulations done on distributive computer networks have been developed to study the folding of the N-terminus of the L9 ribosomal protein using Markov state models approach by Voelz et. al. (26) The fundamental principles of protein folding studied in the works of Shaw and Voelz have extended applications to the sampling of different structural changes related to specific protein function such as flap opening in HIV protease.(27) Signal transduction is another biological process involving proteins that can benefit from studying conformational changes from molecular dynamics simulations; thus, the simulation of the heterotrimeric G-protein is ideal for understanding its transformations during stages of its signal transduction process.

Modeling the solvent is as important as modeling the protein. The energetic contribution of the solvent to binding is very important in understanding protein-protein interactions such as those in the G-protein heterotrimer. The statistical thermodynamics of non covalent binding have been derived for continuum solvation models;(28) these models have become an alternative approach to computationally expensive free energy perturbation approaches to calculate binding free energies in explicit water.(29) In continuum solvation models, the solute is modeled with a dielectric constant that ranges from 1 to 8, and the solvent modeled by a statistical continuum with a dielectric constant of 80.(30) An important continuum solvation method that calculates free energies is the Poisson–Boltzmann (PB) equation,(31) but solutions to this method are computationally expensive, especially for proteins the size of the heterotrimeric G-protein. As a computationally friendly approximation, the generalized-Born (GB) method has emerged as a fast alternative solution to PB methods.(32) Both methods allow for the decomposition of free

energies into residue contributions that help understand free energies of binding in more detail, and help to better understand electrostatic affinity and specificity.

Electrostatic contributions to free energies play a fundamental role in determining specificity of protein-protein interactions. The importance of electrostatics has also been demonstrated for other biological systems (33) and has been applied to the design of biomolecules.(34-35) In this work, an energetic analysis is applied to identify residues with significant electrostatic contributions in the heterotrimeric G-protein that can help understand specificity and its energetic determinants of function that are fundamental in molecular recognition events in signal transduction.(36-37)

1.2.2 Computational studies of G-proteins

Simulations on other G-proteins are helpful in guiding the analysis done in this work. Recent papers on molecular dynamics simulations of other proteins belonging to the Ras GTPase family of proteins include the work by Futsugima and Tsuda exploring the switching mechanism of H-ras gene with a point mutation.(38) Gorfe et. al. simulated 5 GTPase systems for 10 to 15 ns.(39) Three of the systems correspond the α H-ras protein with GDP, GTP and no nucleotide. The other two simulations correspond to an N-ras and a K-ras system without a nucleotide; a mutation in the respective simulations explored the nucleotide switching mechanism that is associated with the oncogenic tendency of these systems to remain in a GTP-like state. A recent publication by Jones et. al. (40) simulated a heterotrimer built from homology modeling based on the 1GP2 systems studied here; their conclusion suggest a connection between the N-terminus alpha helix and switch region two, and the conclusion is supported by experimental observations. Raimondi et. al. (41) explored solvent accessibility of the nucleotide as a factor for its exchange. They also looked at the effect of long range allosteric interactions in small G-proteins as well as in a larger G_{out} for 40 ns each. The simulation of the G-protein in its three states done in this work adds to the growing body of knowledge on this important family of proteins. These three different states correspond to snapshots of the heterotrimeric G-proteins, and their simulations will give insights into the transformations that occur to arrive to these three states. Because the α subunit of the three systems is identical in sequence conformation, differences among the systems can be interpreted as caused by the presence of different nucleotides and the $\beta\gamma$ subunit.

Energetic and backbone structural analysis of the trajectories will be used to characterize the three systems.

1.3 Structural and energetic analysis from trajectories

The following sections give a brief introduction of what will be covered in each of the chapters. In Chapter 2, free energy calculations using generalized-Born from simulations in both explicit and implicit solvent models are compared to experimental values. In Chapter 3, generalized-Born and Poisson–Boltzmann methods are compared for decomposing binding free energies in the G-protein. In Chapter 4, a structural analysis of the G-protein backbone conformations, from its three different states, is performed from extended molecular dynamics simulations. Finally in Chapter 5, an energetic and structural analysis is performed on the three G-protein systems in order to characterize important backbone and side-chain structural differences.

1.3.1 Comparing ensembles generated with implicit and explicit water models help in the interpretation of free energies of binding

The characterization of the energetic landscape of any protein can be done with molecular dynamics simulations in either implicit or explicit solvent. A comparison of these two solvent models is made to experimental binding energies of a family of pyrimidine dicarboxamide bound to MMP-13 protein. The relative simplicity of the systems makes a good case study for comparing generalized-Born dynamics to simulations in explicit water. Generalized-Born free energies of binding are computed for both trajectories. For the ensembles generated with both solvent models, the free energies of binding give good correlation to experiment for the sum of differences in van der Waals and Coulomb contributions, as well as polar and non polar free energy of transfer contributions. The inclusion of entropy in the binding free energies ensembles improves the correlation to experiment for binding energies calculated from explicit solvent, but lowers the correlation for binding energies calculated from implicit solvent. Calculations on these systems confirm that generalized-Born methods give comparable results to explicit solvent in calculating binding free energies. The concepts explained in this paper suggest that entropy gives better results when explicit solvent models are used. In this

chapter, the theoretical foundations necessary for better understanding the results in the next chapter are introduced.

1.3.2 Energetic decomposition with the generalized–Born and Poisson–Boltzmann solvent models: Lessons from association of G-Protein components

Large protein oligomers require refined tools for the understanding of binding free energies. The accuracy of generalized-Born free energy calculations on ensembles generated in explicit water is assessed by comparing them to Poisson–Boltzmann free energy calculations on the heterotrimeric G-protein. Poisson–Boltzmann is considered the gold standard in calculating binding free energies, for which generalized-Born is an approximation. Additionally, both methods allow the decomposition of the binding free energy into the individual contributions from each residue component. Comparing the decomposed energies, as well as overall binding energies, to Poisson–Boltzmann is a more rigorous assessment of the applications of the generalized-Born method. Correlation of total binding free energies, as well as contributions from individual residue types, unveils the advantages and limitations of generalized-Born methods; the chapter concludes that application of this method is best as a filter for identifying relevant components to be calculated with the computationally more demanding and accurate Poisson–Boltzmann method.

1.3.3 Structural motifs of heterotrimeric G-protein subunits and domains from backbone clustering

Long molecular dynamics simulations of the heterotrimeric G-protein in its three different states require methods to group and characterize differences in backbone structures. These differences are due to differences in initial structures from crystals, but more importantly to the presence of different nucleotides and the $\beta\gamma$ subunit. The analysis first aligns all the trajectories in the three systems to rigid sections whose alpha carbons have low B-factor values; this alignment is only applied to the α -subunit, which is identical in sequence in all the systems. Switch regions, shown in Figure 1.2, are of special interest for characterizing α in its three different states. Differences in backbone motions between the GTPase domain and the α -helical domain are analyzed in search of differences that suggest a mechanism of nucleotide exchange. For this purpose, a clustering algorithm is implemented to measure and compare motions in these regions of the proteins. The clustering algorithm is based in Root Mean Square Deviations

between all frames within each simulation, and across frames in different simulations. This methodology shows that motion of the backbone reflects differences due to the presence of GDP, GTP and the $\beta\gamma$ -subunit.

1.3.4 Energetic decomposition of binding and stability of selected components from extended molecular dynamic simulations

This chapter gives a structural and energetic analysis of the heterotrimeric G-protein in its three different states. Because of the length of the calculation, and the sizes of the systems, it is extremely difficult to perform the energetic decomposition on all the components of the protein. To make the analysis practical, the generalized-Born method is used as a preliminary filter for selecting the most relevant components. Poisson–Boltzmann calculations are then performed on components selected from this cutoff. In addition to binding energies of the nucleotides and between α and $\beta\gamma$, folding energies are also considered in the analysis. Folding energies help identify components that are not captured in binding calculations, but that interact with other components that in turn are important for binding affinity and molecular recognition. Folding can also identify inter-subunit interactions in the protein that may not be important in binding but play other functions in the protein. Folding analysis thus provides a more complete energetic picture of heterotrimeric G-protein function. For each of the folding and binding components considered, a detailed description of their structural differences in the three systems is given. In this way, an energetic and structural characterization of the heterotrimeric G-protein gives insights into signal transduction mechanisms.

CHAPTER 2

Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13

This chapter corresponds to a published paper by Carrascal N., and Rizzo R.C. with the same title in Bioorganic & Medicinal Chemistry Letters, 2009. 19(1): p. 47-50. I contributed to this paper by generating the data and analysis of the data. Writing of the text was done by Dr. Robert C. Rizzo with my assistance.

2.1 Abstract

All-atom molecular dynamics (MD) simulations in both explicit and implicit solvent, followed by MM-GBSA energy analysis, have been used to estimate binding free energies of four pyrimidine dicarboxamide inhibitors with human collagenase-3 (MMP-13) for comparison with experimental activities. Energetic analysis reveals that affinity is driven primarily by favorable van der Waals interactions and burial of total surface area. The computed effects of desolvation, as a function of ligand structure, quantitatively show that hydrophilic derivatives pay greater penalties upon binding than their related more hydrophobic analogs.

2.2 Introduction

The initial excitement generated by early matrix metalloproteinase (MMP) inhibitors (MMPIs) for treatment of cancer was dampened by disappointing clinical trial results which showed little or no efficacy.(42-43) In retrospect, early compounds such as marimastat were too broad spectrum and efforts are now focused on design of inhibitors selective for a given MMP. Engel et. al.(44) has recently reported a series of highly selective pyrimidine dicarboxamide inhibitors, shown in Table 2.1, which target human collagenase-3 (MMP-13). These compounds showed no activity against ten other MMPs tested. Uniquely, these are the first reported MMPIs whose mechanism of action is not chelation of the active site catalytic zinc ion.(44)

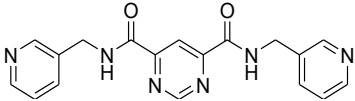
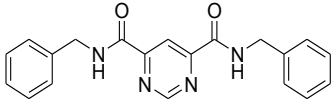
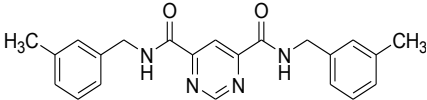
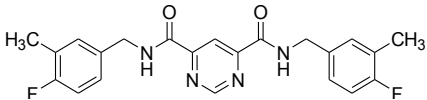
Despite minimal changes to the pyrimidine scaffold, the ligands reported by Engel et. al. (44) span a wide 4 kcal/mol range in binding free energy which makes an ideal test case for evaluating computer-aided design methods. In this report, we have used all-atom molecular dynamics (MD) computer simulations, in conjunction with free energy calculations, in an effort to characterize which specific physical properties modulate binding for this MMPI series. To

evaluate how different water models affect the results, separate MD trajectories for each protein-ligand complex were generated in both explicit TIP3P (45) (TIP3P-MD) and implicit generalized-Born (46) (GB-MD) solvent. Implicit-based MD provides an attractive alternative to explicit solvent simulations when enhanced sampling is desired.

2.3 Theory

For each protein-ligand complex free energies of binding were estimated using the MM-GBSA (47) method. Our laboratory has recently employed this method to quantify inhibition for peptides with HIVgp41 (48) and characterize origins of resistance for small molecules with neuraminidase.(49) An earlier study by Rizzo et. al. (50) used the method to quantify binding and selectivity for the related MMP-1 and MMP-3 systems with good results. Although considered to be an approximate approach, the relatively straightforward setup, ease of use, and the ability to study large structural and conformational changes make use of MM-GBSA an attractive alternative to more computationally intensive methods such as free energy perturbation (FEP).

Table 2.1. Experimental activities of pyrimidine dicarboxamides with MMP-13.

MMPI	Structure	IC ₅₀ ^a	ΔG _{bind} exptl ^b
P01		6600	-7.07
P02		400	-8.73
P03		72	-9.75
P04		8	-11.05

^aIC₅₀ values in nM from reference.(44)
^bΔG_{bind} exptl estimated as RT ln (IC₅₀) in kcal/mol.

Using coordinates saved periodically during an MD simulation of each relevant complex

the binding affinities are estimated via Equation 2.1.

$$\Delta G_{bind} = \Delta E_{coul} + \Delta E_{vdw} + \Delta \Delta G_{hyd} - T\Delta S \quad (2.1)$$

In Equation 2.1, ΔE_{coul} and ΔE_{vdw} are nonbonded protein-ligand intermolecular Coulombic and van der Waals energies, $\Delta \Delta G_{hyd}$ is the change in system hydration (desolvation penalties), and $T\Delta S$ is the change in solute vibrational, rotational, and translational entropy. The energies are computed using a classical molecular mechanics force field. $\Delta \Delta G_{hyd}$ is obtained from $\Delta G_{hyd} = G_{polar} + G_{nonpolar}$ of each species (complex, receptor, and ligand). The GB (46) method is used to estimate G_{polar} and molecular solvent accessible surface (SASA) is used to estimate $G_{nonpolar} = \gamma * SASA + \beta$ using standard constants.(51) $T\Delta S$ energies are computed from normal-mode analysis of energy minimized structures.(47)

2.4 Methods

Setup, simulations, and analysis employed the AMBER8 suite of programs (52) which was used for assignment of force-field parameters (*leap*, *antechamber*), MD simulations/post processing (*sander*, *ptraj*), and normal-mode calculations (*nmode*). AMBER radii (*mbondi2*) with dielectric constants 1 and 78.5 and GB model (*igb=5*) were used for both the original GB-MD simulations and the single-point calculations used to obtain the energy terms in Equation 2.1. For explicit solvent simulations the water was stripped off prior to MM-GBSA analysis.

The catalytic domain of human collagenase-3 (MMP-13) complexed with ligand **P03** (pdb entry 1XUC)(44) was used as a starting point for the simulations as shown schematically in Figure 2.1. Other analogs in Table 2.1 were manually constructed using the coordinates of 1XUC as a guide with the program MOE. (53) His residues were singly protonated at either the epsilon ($N\epsilon$) or delta nitrogen position ($N\delta$) to maximize coordination with zinc. Two zinc and two calcium ions were retained but all crystallographic waters were removed. FF99SB (54) and GAFF (55) parameters were used for the protein residues and ligands respectively. Ligand partial atomic charges were obtained from ChelpG (56) calculations at the HF/6-31G*//HF/6-31G* level of theory using Gaussian98. (57) Zinc parameters were taken from the Stote et. al. (58) nonbonded model.

A multi-step equilibration procedure was used for both simulation protocols with the

intent of obtaining representative and stable ensembles using short simulation times. For explicit solvent (TIP3P-MD), energy minimization for 1000 cycles followed by 15 ps of MD was first performed using a restraint weight of 5.0 kcal/mol/Å² on heavy atoms (steps 1-2). Next, three energy minimizations of 1000 cycles each were done in which heavy atoms restraints were reduced from 2.0, 0.1, to 0.0 kcal/mol/Å² (steps 3-5). Two short 5 ps MD runs followed using weights of 1.0 and 0.5 kcal/mol/Å² (steps 6-7). Finally, a 10 ps MD equilibration followed with 0.5 kcal/mol/Å² restraints only on backbone atoms and ions (step 8). Lastly, equilibration for 505 ps followed by production MD for 500 ps was performed (steps 9-10) in which restraints were used only on ions and backbone atoms further than 10 Å from each ligand (0.5 kcal/mol/Å² weight). All MD runs were at 298.15 K and used a 1 fs time step during equilibration (steps 1-8) and a 2 fs time step during final equilibration and production (steps 9-10). The particle mesh Ewald (PME) (59) method was used with 8.0 Å direct-space nonbonded cutoff. Implicit solvent simulations (GB-MD) employed the exact same multi-step protocol however PME was not used, the nonbonded cutoff was increased to 15.0 Å, and a 1 fs time step was employed.

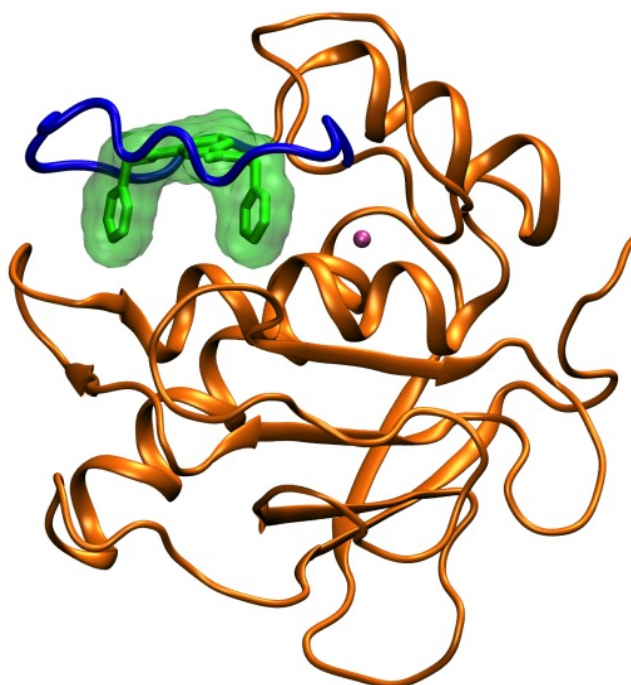


Figure 2.1 MMP-13 structure. MMP-13 (orange) complexed with ligand P03 (green) from pdb entry 1XUC. The catalytic zinc ion is shown in cyan and the flexible binding site loop is shown in blue.

2.5 Results

All simulations were well-behaved as measured by root-mean-square stability and plots of energy vs. time. Energetic results from the last 500 ps are shown in Table 2.2. Here, free energies of binding, calculated both with ($\Delta G_{\text{MM-GBSA+E}}$) and without ($\Delta G_{\text{MM-GBSA}}$) solute entropy are listed along with a breakdown of the individual energy terms in eq 1. Both explicit and implicit-solvent results are presented along with correlation coefficients (R^2 values) computed for all columns with the experimental activities. However, given that the dataset contains a small number of ligands, these R^2 values are useful only in the context of viewing general trends. Nevertheless, an inspection of the correlations, along with the raw energy components, clearly reveal that both simulation protocols yield the same trends with the sole exception being TΔS as discussed further below. Figure 2.2 highlights the overall good agreement.

Table 2.2 Contributions towards calculated free energies of binding for pyrimidine dicarboxamide inhibitors with MMP-13.^a

MM PI	ΔE_{vdw} (N=501) A	ΔE_{coul} (N=501) B	$\Delta \Delta G_{\text{polar}}$ (N=501) C	$\Delta \Delta G_{\text{nonpolar}}$ (N=501) D	TΔS (N=21) E	$\Delta G_{\text{MM-GBSA}}$ A+B+C+D G	$\Delta G_{\text{MM-GBSA+E}}$ A+B+C+D+E H	ΔG_{bind} exptl ^b I
Explicit Water (TIP3P-MD)								
P01	-52.65 ± 0.11	-22.97 ± 0.20	44.77 ± 0.15	-5.63 ± 0.00	-18.86 ± 1.93	-36.48 ± 0.13	-17.61 ± 1.93	-7.07
P02	-52.07 ± 0.12	-17.35 ± 0.13	35.72 ± 0.10	-5.76 ± 0.00	-20.28 ± 1.45	-39.44 ± 0.12	-19.17 ± 1.46	-8.73
P03	-59.42 ± 0.11	-21.26 ± 0.14	40.22 ± 0.10	-6.09 ± 0.00	-26.16 ± 2.00	-46.55 ± 0.12	-20.38 ± 2.00	-9.75
P04	-60.30 ± 0.11	-25.29 ± 0.14	44.59 ± 0.10	-6.35 ± 0.00	-26.57 ± 1.90	-47.35 ± 0.12	-20.79 ± 1.90	-11.05
	$R^2 = 0.74$	$R^2 = 0.12$	$R^2 = 0.00$	$R^2 = 0.94$	$R^2 = 0.85$	$R^2 = 0.89$	$R^2 = 0.96$	
Implicit Water (GB-MD)								
P01	-52.47 ± 0.14	-25.74 ± 0.22	47.04 ± 0.17	-5.80 ± 0.01	-15.85 ± 2.20	-36.97 ± 0.14	-21.12 ± 2.21	-7.07
P02	-54.81 ± 0.12	-21.13 ± 0.16	39.70 ± 0.13	-5.91 ± 0.01	-11.36 ± 1.76	-42.14 ± 0.13	-30.78 ± 1.77	-8.73
P03	-57.76 ± 0.15	-18.72 ± 0.16	38.94 ± 0.11	-6.46 ± 0.01	-21.06 ± 2.01	-43.99 ± 0.15	-22.93 ± 2.02	-9.75
P04	-59.62 ± 0.14	-24.76 ± 0.17	44.90 ± 0.14	-6.52 ± 0.01	-18.43 ± 1.58	-46.00 ± 0.14	-27.57 ± 1.59	-11.05
	$R^2 = 0.98$	$R^2 = 0.06$	$R^2 = 0.09$	$R^2 = 0.85$	$R^2 = 0.23$	$R^2 = 0.97$	$R^2 = 0.18$	
^a All energies ± standard error of the mean in kcal/mol. ^b ΔG_{bind} exptl from Table 2.1.								

Notably, both MD simulation protocols yield comparable binding free energies (Table 2.2 column G, $\Delta G_{\text{MM-GBSA}}$) which correctly reproduce the experimental ordering provided TΔS is omitted. In both cases, increased favorable intermolecular packing, as embodied in ΔE_{vdw} and $\Delta \Delta G_{\text{nonpolar}}$ (Table 2.2 columns A and D) appears to best explain the experimental variation. In general, more van der Waals contact occurs and more surface area is systematically buried as additional atoms are added to the ligand scaffold.

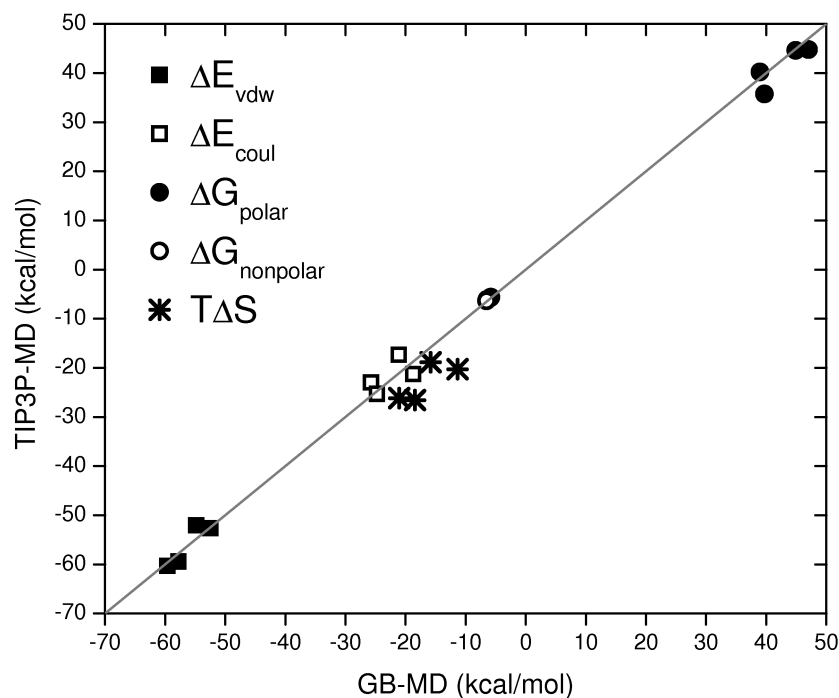


Figure 2.2. Comparison between energy components obtained from explicit solvent (TIP3P-MD) and implicit solvent (GB-MD) trajectories.

Both simulation protocols also yield the same trends in intermolecular Coulombic interactions (Table 2.2 column B). The ΔE_{coul} terms are consistently more favorable for the more polar **P01/P04** pair versus the more hydrophobic **P02/P03** pair. Further, the simulations quantify the important desolvation effects which are a function of ligand structure. The more hydrophilic **P01** (pyridine) and **P04** (o-fluoro toluene) derivatives show greater desolvation penalties (Table 2.2, $\Delta\Delta G_{\text{polar}}$ column C) of ca. 45 – 47 kcal/mol compared with the more hydrophobic **P02** (benzene) and **P03** (toluene) compounds at 36 – 40 kcal/mol. Consistent with these quantitative GB results are the visual patterns of hydration observed in snapshots from the explicit solvent MD simulations (Figure 2.3). In all cases, water is clustered about the central polar dicarboxamide pyrimidine scaffold. However, for the more polar derivatives, additional waters cluster about the terminal pyridine (**P01**) and o-fluoro toluene (**P04**) groups in comparison with the more hydrophobic benzene (**P02**) and toluene (**P03**) groups.

Compared with the other energy terms (Table 2.2, Figure 2.2) estimates for $T\Delta S$ appear to be more sensitive to the sampling protocol used to derive the ensembles. Explicit MD trajectories yield $T\Delta S$ energies which lead to a systematic increase in entropy across the ligand series which also correlate with the experimental ΔG_{bind} . The implicit-derived results show a

different pattern. Surprisingly, an examination of individual results from snapshots which comprise the average entropy values listed in Table 2.2 reveal that some GB-derived terms yield positive values.

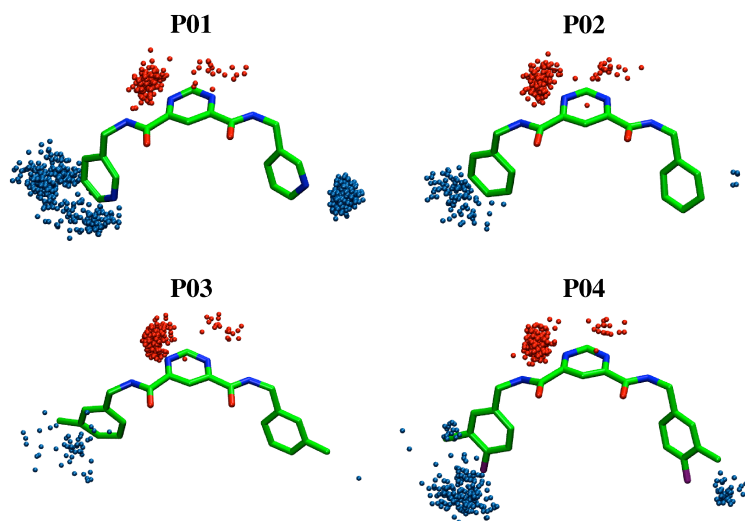


Figure 2.3. Solvation patterns for MMPIs from explicit solvent TIP3P-MD trajectories. Spheres indicate water oxygens within 2.5 Å of ligand heavy atoms for the center pyrimidine ring (red) or outer rings (blue). (N=500).

Positive $T\Delta S$ terms are not intuitive as the overall change in free energy is expected to decrease ($\Delta G = \Delta H - T\Delta S$) when solute entropy changes are included. In contrast, explicit solvent calculations showed no such anomalies. Here, the greater variation inherent in GB vs. TIP3P ensembles is likely the dominant factor for the observed $T\Delta S$ differences. As illustrated in Figure 2.4, coordinates from implicit MD show greater variation than the corresponding set from explicit MD after the extensive energy minimizations required for the nmode calculations used to compute $T\Delta S$. In general, use of larger ensembles is probably more important for entropy estimates when using implicit simulation results given the greater variation observed here for minimized GB-MD snapshots. Studies to more fully address how sampling influences $T\Delta S$ results, across different simulations protocols, are underway in our laboratory.

Incorporating the intuitively correct $T\Delta S$ results derived from explicit solvent simulations yields a strong correlation with experimental ΔG_{bind} of $R^2=0.96$ (Table 2.2 column H). Most striking is the fact that the computed *relative* free energies of binding ($\Delta\Delta G_{\text{bind}}$) yield

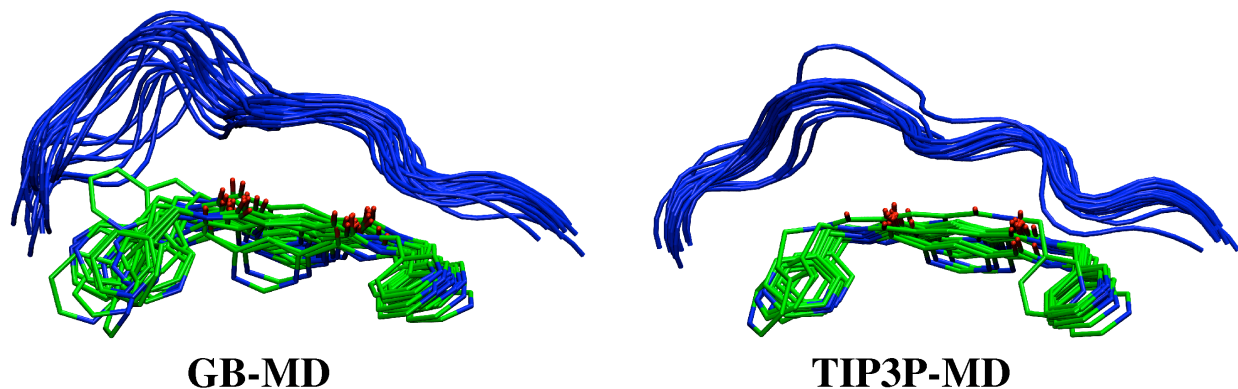


Figure 2.4. Ensemble differences from implicit and explicit solvent models. Energy minimized snapshots (N=21), from both simulation protocols, for the P01-MMP13 complex. Binding site loop in blue.

quantitative agreement with the experimental results as shown in Table 2.3. With high accuracy, the explicit $\Delta\Delta G_{\text{MM-GBSA+E}}$ results quantify the effects of functional group substitution for changing pyridines on **P01** to benzene (-1.66 exptl vs -1.56 calcd, **P02**), adding ortho methyl groups (-2.68 exptl vs -2.77 calcd, **P03**), and adding para fluorines (-3.88 exptl vs -3.18 calcd, **P04**). Importantly, this accuracy approaches that of FEP methods which are historically regarded as the gold standard in binding energy calculations with errors on the order of only ca. 0.25 to 0.5 kcal/mol. Further investigation of MM-GBSA methods is clearly warranted.

Table 2.3. Relative free energies of binding ($\Delta\Delta G$) from TIP3P-MD simulations for pyrimidine dicarboxamide inhibitors with MMP-13.^a

MMPI	$\Delta\Delta G_{\text{bind exptl}}^a$	$\Delta\Delta G_{\text{MM-GBSA+E}}$
P01	0.00	0.00
P02	-1.66	-1.56
P03	-2.68	-2.77
P04	-3.98	-3.18

^aExperimental and predicted values from Table 2.1 normalized to P01.

2.6 Conclusion

In summary, results from all-atom simulations of four pyrimidine dicarboxamide inhibitors with MMP-13 have been used to characterize what drives binding and to test the effects of using explicit vs implicit solvent MD. Both modeling methods reveal that variations in van der Waals interactions (ΔE_{vdw}) and burial of surface area ($\Delta\Delta G_{\text{nonpolar}}$) best describe the experimental results. The calculations also yield physically sound $\Delta\Delta G_{\text{polar}}$ and ΔE_{coul} energies with the more polar compounds showing enhanced Coulombic interactions in the binding site

and an associated larger desolvation penalty. An examination of explicit solvent results reveals local changes in hydration, as a function of ligand structure, in good visual agreement with the GBSA results.

The strong correspondence in the results suggest that overall both MD protocols sample comparable regions of the energy landscapes which is encouraging for use of GB-derived ensembles for estimation of ΔG_{bind} . The exception is solute entropy for which implicit and explicit-derived results yield T Δ S terms which are not correlated. The addition of implicit-derived entropy to the calculated free energies of binding yield diminished agreement with experiment. However, for explicit solvent results good accord is obtained in all cases. In particular, relative free energies of binding ($\Delta\Delta G_{\text{bind}}$) are in striking quantitative agreement. Future studies should examine the use of larger datasets, longer simulations, and in particular larger ensembles for estimation of T Δ S to more fully assess convergence.

CHAPTER 3

Energetic decomposition with the generalized-Born and Poisson–Boltzmann solvent models: lessons from association of G-protein components

This chapter corresponds to a published paper by Carrascal N., and Green D.F. with the same title in the Journal of Physical Chemistry B, 2010. 114(15): p. 5096-5116. I contributed to this paper by generating the methods for generating the data and plots. Analysis and writing of the final text was done by Dr. David. F. Green with my assistance.

3.1 Abstract

Continuum electrostatic models have been shown to be powerful tools in providing insight into the energetic of biomolecular processes. While the Poisson–Boltzmann (PB) equation provides a theoretically rigorous approach to computing electrostatic free energies of solution in such a model, computational cost makes its use for large ensembles of states impractical. The generalized-Born (GB) approximation provides a much faster alternative, although with a weaker theoretical framework. While much attention has been given to how GB recapitulates PB energetics for the overall stability of a biomolecule or the affinity of a complex, little attention has been given to how the contributions of individual functional groups are captured by the two methods. Accurately capturing these individual electrostatic components is essential both for the development of a mechanistic understanding of biomolecular processes and for the design of variant sequences and structures with desired properties. Here, we present a detailed comparison of the group-wise decomposition of both PB and GB electrostatic free energies of binding, using association of various components of the heterotrimeric-G-protein complex as a model. We find that, while net binding free energies are strongly correlated in the two models, the correlations of individual group contributions are highly variable; in some cases, strong correlation is seen, while in others, there is essentially none. Structurally, the GB model seems to capture the magnitude of direct, short-range electrostatic interactions quite well but performs more poorly with moderate-range “action-at-a-distance” interactions. GB has a tendency to overestimate solvent screening over moderate distances, and to underestimate the costs of desolvating charged groups somewhat removed from the binding interface. Despite this, however, GB does seem to be quite effective as a predictor of those groups that will be computed to be most significant in a PB-based model.

3.2 Introduction

The interactions between biological molecules lie at the heart of the majority of fundamental biological processes. As a result, a deep understanding of these interactions is an essential component of understanding biology as a whole. The mechanisms by which extracellular signals are transduced into intracellular changes in cellular behavior are a perfect example of this, being driven by cascades of interactions between various signaling molecules.

One of the prototypical models for signal transduction is the heterotrimeric-G-protein signal transduction pathway.^(5, 8, 15) Heptahelical, integral membrane proteins (the G-protein-coupled receptors, GPCRs) interact with a diverse range of extracellular ligands. Intracellularly, the GPCRs interact with heterotrimeric G-proteins, which consist of three subunits (α , β , and γ). In the inactive state, the G-proteins are in a trimeric form, with the α -subunit additionally associated with guanosine-5'-diphosphate (GDP). Upon receptor activation (by ligand binding), the receptor promotes exchange of the GDP bound to $G\alpha$ with guanosine-5'-triphosphate (GTP); this leads to a conformational change in $G\alpha$ that results in dissociation from the β - and γ -subunits (which remain complexed). Both $G\alpha$ and $G_{\beta\gamma}$ are then free to interact with additional cellular-signaling proteins; different variants of the G-protein subunits have different secondary targets.

In mammalian genomes, there are multiple variants of each subunit—at least 20 α (16 well-characterized, and several putative forms that are less or uncharacterized), 7 β (including multiple isoforms from individual genes), and 12 γ variants;⁽⁶⁰⁾ it is the appropriate combination of these which leads to the correct coupling of a specific GPCR to the cognate cellular response. While a great deal of work has been done to decode the specific interactions made between various G-protein subunits, there remain many open questions regarding which combinations of subunits are possible, as well as which are physiologically relevant.⁽⁶¹⁻⁶⁵⁾ A detailed understanding of the energetics underlying subunit association could provide important insight into this problem.

The association of biological molecules typically occurs in an aqueous environment of moderate ionic strength; thus, solute–solvent interactions must be appropriately treated in any energetic model. Fully explicit, all-atom simulations with the biomolecule embedded in a box of water molecules and ions, with periodic-boundary conditions to reduce boundary effects, have been very successful in providing a deep understanding of biomolecular structure and dynamics

in solution.(54, 66-67) However, while well-suited to determining the dynamic, configurational ensembles of a single species (or complex), many challenges arise in computing free-energy differences between the bound and unbound states of a complex. In particular, accurate estimations of free energies in explicit solvent require adequate sampling of solvent degrees of freedom; additional difficulties arise in comparing two independent simulations where the number of explicit solvent molecules may vary. Accurate calculations of free-energy differences in solution can be achieved though free-energy perturbation techniques, but these approaches are tremendously costly.(28, 68)

It has long been recognized that a powerful alternative to explicit-solvent free-energy perturbation approaches to computing solute-solvent interaction free energies is the use of continuum solvent models.(29, 69) Representing the interior of a biomolecule as a region of low dielectric (typically between 1 and 8) and solvent as a region of high dielectric (roughly 80 for water at room temperature) with a Debye–Hückel-like treatment of mobile ions, the electrostatic potential of a molecule, and thus its free energy, can be computed by solution of the Poisson–Boltzmann (PB) equation.(31, 70) PB-based continuum solvent models have been used to provide important insights into many aspects of molecular recognition: understanding the optimum of enzymes for substrate or transition-state binding;(71-73) dissecting the contribution of various groups to binding affinity;(33, 74) predicting the relative affinities of ligands;(75-77) understanding protein-membrane association;(78) and applications to the design of protein complexes.(35, 79)

Over the past decade, there has been an increasing interest in faster alternatives to PB-based models for use in applications that require many repeated evaluations, such as high-throughput virtual screening of potential drug molecules and molecular dynamics simulations. One of the most commonly used of these is the generalized-Born (GB) model originally proposed by Still and co-workers.(46, 80-81) Several implementations of the GB model have been applied to many problems with a great deal of success, but some concerns have also been raised.(82-86) For example, Brooks and co-workers have noted an imbalance between solute-solvent and intersolute interactions,(87) and Onufriev et al. have found that GB-based methods can be sensitive to the method by which effective Born radii are computed.(88)

One of the great advantages of Poisson–Boltzmann-based models is that the free-energies are, in the context of the linearized-PB equation, pairwise decomposable in terms of individual partial atomic charges. This allows for a rigorous decomposition of contributions from individual chemical groups, such as the amino-acid side chains of a protein; such a detailed understanding of the role of particular residues in contributing to protein stability and complex affinity has been shown to be an important tool in the engineering of proteins with enhanced properties.⁽⁸⁹⁻⁹¹⁾ Additionally, the pairwise decomposability of PB-based electrostatics has led to an elegant optimization theory, through which the electrostatic interactions in a complex may be tuned to provide a provably optimal contribution to binding.^(72, 92-93)

Generalized-Born models are equally pairwise decomposable, and thus should be applicable in the same framework that has made PB-based analysis so powerful. However, the ability of GB-based models to recapitulate PB-based energetics at this level of detail has not been extensively explored. Here, we present results from the application of both PB- and GB-based component analysis to the interactions between various components of the heterotrimeric G-proteins. The results highlight important differences that must be taken into consideration in understanding any energetic decomposition with GB models. We begin with a review of the underlying theory, and then present the computational results along with a detailed discussion.

3.3 Theory

3.3.1 The linearized-Poisson–Boltzmann model is pairwise decomposable

Implicit solvent models replace an atomistic description of solvent with a continuum model; typically, this solvent continuum is described by macroscopic solvent properties such as a dielectric constant and surface tension. While debate remains as to the most appropriate treatment of nonpolar effects, the physical theory for interactions of charged particles in a (nonuniform) dielectric continuum is well-defined.⁽⁹⁴⁾

Given an arbitrary density of charge, $\rho(\vec{r})$, in a medium of spatially varying dielectric, $\epsilon(\vec{r})$, the potential in all space, $\phi(\vec{r})$, is determined by the solution to the Poisson equation, $\nabla \cdot \epsilon(\vec{r})\nabla\phi(\vec{r}) = -4\pi\rho(\vec{r})$. When monovalent mobile ions are additionally considered (with a Debye–Hückel-like treatment), this leads to the Poisson–Boltzmann (PB) equation:

$$\nabla \cdot \boldsymbol{\varepsilon}(\vec{r})\nabla \phi(\vec{r}) + \kappa^2(\vec{r}) \sinh \phi(\vec{r}) = -4\pi\rho(\vec{r}) \quad (3.1)$$

The Poisson equation simply lacks the second term, as $\kappa^2(\vec{r}) = 0, \nabla\vec{r}$ if no mobile ions are present, and a well-defined extension to multivalent ions exists.(95) When the electrostatic potential in solution is low, the hyperbolic sine term may be approximated by the leading (linear) term in the power series expansion, giving the linearized-Poisson–Boltzmann (LPB) equation:

$$\nabla \cdot \boldsymbol{\varepsilon}(\vec{r})\nabla \phi(\vec{r}) + \kappa^2(\vec{r})\phi(\vec{r}) = -4\pi\rho(\vec{r}) \quad (3.2)$$

The PB equation, both in its linearized and nonlinear forms, can be solved numerically through many methods, including finite-difference,(70, 96) finite element,(97) and boundary element (98) implementations.

Decompositions of LPB energetics have been applied to many systems, but the theory underlying the decomposition is worth reviewing. Given the potential in all space, the free energy of a state in the LPB model is given by

$$G = \frac{1}{2} \int_{\mathcal{V}} \rho(\vec{r})\phi(\vec{r})dV \quad (3.3)$$

with the integral taken over all space. This is indeed a free energy, as the cost of creating reaction field in a dielectric medium (including entropic terms) is included. In a typical application involving biological macromolecules, the molecular charge distribution is approximated by a set of point charges at atomic centers. Thus, the term $\rho(\vec{r})$ vanishes at all points other than atomic centers, and the integral in Equation 3.3 becomes a simple sum over atoms:

$$G = \frac{1}{2} \sum_{j=1}^N q_i \phi_i \quad (3.4)$$

where q_i is the partial atomic charge on atom i and ϕ_i is the total electrostatic potential at atom i . Now, since the LPB equation describes a linear-response model, the potential at a given point can be described by a linear combination of the potentials due to all atoms in the system.

Furthermore, the potential due to a given atom is the potential due to a unit charge at that position, ϕ'_{ij} , scaled by the partial atomic charge, q_j :

$$\phi_i = \sum_{j=1}^N q_j \phi'_{ij} \quad (3.5)$$

Combining Equations 3.4 and 3.5, we get the result

$$G = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N q_i \phi'_{ij} q_j \quad (3.6)$$

Thus, we see that the electrostatic free energy of a system described by the LPB model is pairwise decomposable by atom. This feature allows the energy to be similarly decomposed into pairwise terms involving various sets of atoms, such as chemical functionalities or individual amino acids in a protein. We may simplify this expression by defining the electrostatic potential matrix, $\Phi = 1/2[\phi'_{ij}]$, and a partial atomic charge vector, $\vec{Q} = [q_i]$, which gives the matrix expression for the electrostatic free energy:

$$G = \vec{Q}^T \Phi \vec{Q} \quad (3.7)$$

It is worthwhile to briefly discuss the origin of the factor of one half in the free-energy expression, which has two distinct sources. The off-diagonal terms of the Φ matrix (pairwise interactions) are halved in order to prevent double counting; in the double-sum form of Equation 3.6, the factor of half may be dropped if the second sum is taken over $j > i$ rather than over all j . These terms involve both a Coulombic term and a term due to dielectric screening of the Coulombic interaction. The diagonal terms of Φ (self-energies) are halved for a very different reason. These terms, which correspond to the interaction of a single charge with its own reaction field, are not counted twice in the summation. However, the energetic cost of creating the reaction field is equal to half of the total interaction energy of the charge with the reaction field but of opposing sign. Thus, the factor of one half for the diagonal elements accounts for the cost of reorganizing the dielectric medium; as this includes both entropic and enthalpic terms, this leads to a free energy, and not simply a potential energy.

Note that, in the full nonlinear PB model, the energy is not simply given by Equation 3.3, as an additional integral over all space is required to account for the energy of the mobile ions. Additionally, the potential in the nonlinear form is not a linear superposition of potentials from unit charges. Thus, this decomposition strictly applies only to the linearized form.

3.3.2 Free-energy differences in the linearized-Poisson–Boltzmann model

The atom-by-atom pairwise decomposability of the LPB model for the free energy of a single state leads directly to similar pairwise decomposability for differences in free energy, including the free energies of solvation, of molecular association, and of conformational changes (such as protein folding). Given two states, A and B, with corresponding potential matrices Φ_A and Φ_B , the electrostatic free-energy change going from A to B is

$$\begin{aligned}\Delta G &= \vec{Q}^T \Phi_B \vec{Q} - \vec{Q}^T \Phi_A \vec{Q} \\ \Delta G &= \vec{Q}^T \Delta \Phi_{AB} \vec{Q}\end{aligned}\tag{3.8}$$

where $\Delta \Phi_{AB} = (\Phi_B - \Phi_A)$ is the difference of the two potential matrices. Note that this presumes a nonpolarizable description of the molecular system, where partial atomic charges do not vary with state. In a polarizable model, the pairwise decomposability remains, but the representation of Equation 3.8 does not hold.

For a solvation free energy, the two states correspond to the same molecular configuration in two different dielectric environments; most often, the external dielectric environment is that of aqueous solution in one state and that of vacuum ($\epsilon = 1$) in the other. As a result, the Coulombic terms (in a vacuum) are identical in the two states, and the difference matrix consists only of differences in interactions with the reaction field; in moving from a higher to lower external dielectric, the difference matrix will then be positive definite.⁽⁹⁹⁾ For a conformational free-energy change, the molecular composition and general dielectric environment remain fixed, but the positions of both partial atomic charges and the boundary between dielectric regions vary. Thus, the difference matrix will contain both Coulombic and reaction-field components, and no general statements can be made about the properties of the matrix. When the conformational change of interest is that of protein folding, often an approximation to the unfolded state is made in which short, contiguous sequences of amino acids (1-3) are modeled as isolated molecules. In this model, the same formalism holds, and the

potential matrix for the unfolded state will be block diagonal, with zero elements for all atoms not present in the same reference molecule and each diagonal block corresponding to the potential matrix for the model. Thus, in the difference matrix, only the diagonal blocks will differ from the potential matrix of the folded state. For a binding free energy, the two states consist of (A) a complex and (B) each component of the complex in isolation. The unbound-state matrix, similar to the unfolded-state matrix discussed above, will be block diagonal, with all intermolecular elements zero; each diagonal block will be the potential matrix of the isolated molecule. Again, this leads to the result that the off-diagonal blocks of the difference matrix are unchanged from that of the complex potential matrix. In a rigid-body binding model, where the unbound states are identical in conformation to the bound state, the Coulombic entries of the diagonal blocks are identical in both states, and thus the diagonal blocks of the difference matrix contain only the difference of the reaction field terms.

3.3.3 Energetic decomposition in the linearized-Poisson–Boltzmann model

The form of Equation 3.8 makes it clear that the differences in electrostatic free energies are strictly pairwise decomposable over the atoms of the system, given Φ_A and Φ_B . These matrices do not depend on the magnitudes of the charges in the system but do depend both on the position of the atomic charges and on the boundary separating high from low dielectric (and nonzero ionic strength from zero). The effect of the dielectric boundary on the potential matrices is fundamentally not pairwise decomposable, and this must be considered in the definition of an appropriate reference state. Thus, we raise the concept of a hydrophobic isostere—a hypothetical molecule with an identical shape as that of interest but with some subset of partial atomic charges set to zero. Electrostatic free-energy differences in such a system will be represented by an identical $\Delta\Phi$ matrix, but the charge vector will have the appropriate subset of charges zeroed. We define zero energy to be that of the fully hydrophobic system ($\vec{Q} = \vec{0}$).

The atomic-level pairwise decomposability of the model allows for pairwise decomposability across any groupings of atoms. Consider a set of m groups, each of which may have its charges defined by a vector \vec{q}_i . Then,

$$\vec{Q}^T = [\vec{q}_1^T \vec{q}_2^T \dots \vec{q}_m^T] \quad (3.9)$$

where we recall that the ordering of the indices for the charge vector and the potential matrix is arbitrary, as long as it is consistent. We may similarly consider $\Delta\Phi$ as a block matrix, divided in the same way:

$$\Delta\Phi = \begin{bmatrix} \Delta Q_{11} & \Delta Q_{12} & \cdots & \Delta Q_{1m} \\ \Delta Q_{21} & \Delta Q_{22} & \cdots & \Delta Q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta Q_{m1} & \Delta Q_{m2} & \cdots & \Delta Q_{mm} \end{bmatrix} \quad (3.10)$$

The free-energy expression then becomes

$$\Delta G = \sum_{i=1}^m \sum_{j=1}^m \vec{q}_i^T \Delta\Phi_{ij} q_j \quad (3.11)$$

The meaning of each term in Equation 3.11 depends specifically on how the groups were defined and on the physical process being considered. Consider, for example, a binding reaction between two species (L and R), where the groups are defined simply as all the atoms in each species. In this case, we have

$$\Delta G = \vec{q}_L^T \Delta\phi_{LL} \vec{q}_L + \vec{q}_L^T \Delta\phi_{LR} \vec{q}_R + \vec{q}_R^T \Delta\phi_{RL} \vec{q}_L + \vec{q}_R^T \Delta\phi_{RR} \vec{q}_R \quad (3.12)$$

The terms $\vec{q}_L^T \Delta\phi_{LL} \vec{q}_L$ and $\vec{q}_R^T \Delta\phi_{RR} \vec{q}_R$ correspond to the difference in the interactions of an individual species with its own reaction field in the bound and unbound states, which may be termed the desolvation energy for each species. As $\Delta\Phi$ is symmetric, the two off-diagonal terms are equal, and their sum $2\vec{q}_L^T \Delta\phi_{LR} \vec{q}_R$ corresponds to the solvent-screened Coulombic interaction between the two species in the bound state; the factor of 2 simply accounts for the factor of one half that was included in our definition of $\Delta\Phi$. Thus, we have the result

$$\Delta G = \Delta G_L^{des} + \Delta G_R^{des} + \Delta G_{LR}^{int} \quad (3.13)$$

where the desolvation terms are quadratic in the charges of each species and the interaction term involves a linear contribution from each. This formalization allows for an elegant technique for finding a charge distribution on one species that optimizes the binding free energy to its partner; details of this procedure and applications have been extensively discussed in the literature.(71-73, 92-93, 99-102)

Another common decomposition is into the individual chemical functionalities of a molecule. For a protein, each amino acid is easily divided into three groups: one containing all atoms of the side chain, another containing the backbone amino group, and a third containing the backbone carbonyl. Again considering a binding reaction, we obtain terms of three types (unrelated to the division of each amino acid into three groups). These are the following:

- The diagonal terms, $\vec{q}_i^T \Delta\phi_{ii} \vec{q}_i$, which correspond to the difference (between the bound and unbound state) of the interactions that the i th group makes with its own reaction field. This is termed the desolvation energy of group i , analogously to the desolvation of species A and B, above.
- The off-diagonal terms, $\vec{q}_i^T \Delta\phi_{ij} \vec{q}_j$, where group i and group j belong to different binding partners. These are easily interpreted as the solvent-screened Coulombic interactions between the two groups in the bound state, analogous to the interaction term above, and are termed direct interactions.
- The off-diagonal terms, $\vec{q}_i^T \Delta\phi_{ij} \vec{q}_j$, where group i and group j belong to the same binding partner. In this case, there are nonzero contributions from both the bound and unbound states, and in the rigid-body binding model, the Coulombic terms in the two states perfectly cancel. Thus, these terms correspond to the difference in the solvent screening that intramolecular interactions experience in the bound and unbound states, referred to as *indirect interactions*.

To compute the $\Delta\Phi$ matrix, a separate solution to the LPB equation must be found for each atom in the system, in each of the two states. For each state, considering a system in which a single atom is charged to 1 (and all others are replaced with their hydrophobic isosteres) gives the potential at all atoms due to a unit charge at one specific atom. The vector of these potentials forms a single row of the Φ matrix for that state.

However, while the full $\Delta\Phi$ matrix provides the motivation for energetic decomposition, it is generally not computed; rather, a single calculation can be done for each group in the blockmatrix representation. A system in which a single group is charged (with appropriate partial atomic charges) and all others are replaced with hydrophobic isosteres provides the potential due to that single group; this provides each of the terms from Equation 3.11 ($\vec{q}_i^T \Delta\phi_{ij} \vec{q}_j$) for a

particular group, i . Details and applications of these decompositions have also been discussed in the literature.(33, 89, 103)

3.3.4 Energetic decomposition in the generalized-Born model

As discussed above, in the linearized-PB model, a separate calculation must be done for each group. For a full residue-by residue decomposition of a large protein complex, this can involve several thousand individual calculations. While this is entirely feasible for a single structure, it becomes computationally costly to take this approach to ensembles of structures of any significant size. For example, to perform a complete analysis over an ensemble of 1000 snapshots from a constant temperature molecular dynamics simulation could require millions of solutions of the LPB equation.

The generalized-Born model has gained a great deal of favor in providing a fast alternative to Poisson–Boltzmann electrostatics in situations where speed is a key factor, such as implicit-solvent molecular dynamics and high-throughput virtual screening. Thus, the question naturally arises as to whether GB may provide a useful approximation to PB-based energetic decomposition. In the GB model, each atom in the system is described by an effective Born radius, which is the radius for which a sphere with a centrally located charge would have the same solvation free energy as the charge located at the atomic center in the actual system of interest; for unit charges, these values correspond to the diagonal elements of the Φ matrices discussed above. Given the effective Born radii, the free energy is approximated by the Still equation:(46)

$$G = \frac{1}{8\pi\epsilon_0} \left[\sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{\epsilon_{\text{int}} \|\vec{r}_{ij}\|} - \left(\frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{ext}}} \right) \frac{q_i q_j}{\sqrt{\|\vec{r}_{ij}\|^2 + R_i R_j \exp(-\|\vec{r}\|^2 / 4R_i R_j)}} \right] \quad (3.14)$$

where the first term (skipped for $i = j$) corresponds to Coulomb's Law in uniform dielectric, ϵ_{int} , and the second term accounts for solvent polarization (solvent dielectric, ϵ_{ext}). Clearly, this expression can be expressed in an analogous fashion to Equation 3.7, with

$\Phi_{ij} = (1/\epsilon_{\text{int}} \|\vec{r}_{ij}\|) - [(1/\epsilon_{\text{int}}) - (1/\epsilon_{\text{ext}})] \{1 / (\|\vec{r}_{ij}\|^2 + R_i R_j \exp(-\|\vec{r}\|^2 / 4R_i R_j))\}^{1/2}$ Thus, the decompositions with respect to groups of charges apply equally well to GB-based electrostatics.

For the self-energy of individual charges, the Still equation is rigorously correct when provided exact effective Born radii; the pairwise interactions converge to the energy of a dipole in a sphere in the small distance limit and to the interaction of two isolated point charges in the far distance limit. In practice, however, exact effective Born radii are not known but rather are approximated by a number of different approaches.(80, 83, 104-105)

Case and co-workers have also derived an extension to the GB model that allows for a treatment of mobile ions;(106) the extension simply involves a substitution in Equation 3.14 of

$$\left(\frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{ext}}} \right) \rightarrow \left(\frac{1}{\epsilon_{\text{int}}} \frac{e^{-\kappa \sqrt{\|\vec{r}_{ij}\|^2 + R_i R_j}} \exp(-\|\vec{r}\|^2 / 4 R_i R_j)}{\epsilon_{\text{ext}}} \right) \quad (3.15)$$

Note that, in the GB formulation of the potential matrix Φ , only the effective Born radii are needed; this is a result of the fact that Equation 3.7 separates contributions from geometry and from partial atomic charge. However, whereas in a PB-based model a full solution of the Poisson–Boltzmann equation is required for each row in the potential matrix, a single evaluation of the GB energy (in any implementation) yields all effective Born radii, and thus one energetic evaluation provides the full potential matrix. This could potentially result in tremendous savings in computational cost; while an individual GB computation may be 10- to 100-fold faster than a single solution of the PB equation, when upward of a thousand components are involved, the savings becomes 4 or 5 orders of magnitude.

However, while the potential savings in computation cost are exceptional, it is important to consider accuracy as well. While many studies have considered the overall ability of generalized-Born methods to recapitulate Poisson–Boltzmann-based solvation and binding free energies, there have been few studies of how the two methods compare when energetics are decomposed into individual contributions. The remaining sections of the manuscript discuss these issues.

3.4 Methods

3.4.1 Structure preparation

Initial heavy-atom coordinates were obtained from the Protein data Bank (1gia, 1gdd, and 1gp2).⁽¹⁰⁷⁾ The positions of heavy atoms with missing electron density were modeled using coordinates from the other structures (where these atoms were observed). Protonation states were chosen with the REDUCE program, as were the orientations of amides and imidazole rings;⁽¹⁰⁸⁾ this resulted in a single protonated histidine in the subunit of the trimer (1gp2), with all other histidines in a neutral state, all aspartates and glutamates negatively charged, and all lysines and arginines positive. Hydrogen atom positions were determined with the HBUILD module ⁽¹⁰⁹⁾ of the CHARMM computer program.

3.4.2 Molecular dynamics

All-atom, explicit solvent molecular dynamics simulations were performed using the CHARMM ⁽¹¹⁰⁾ and NAMD ⁽¹¹¹⁾ programs with PARAM22/27 parameters;⁽¹¹²⁻¹¹³⁾ system setup and postprocessing were done with CHARMM, while production dynamics simulations were run using NAMD. Each complex was placed in a pre-equilibrated box of TIP3P waters,⁽⁴⁵⁾ with a minimum of 10 Å between any solute atom and the box edge in all directions. Random water positions were replaced with enough sodium and chloride ions to yield physiological (145 mM) ionic strength; the sodium to chloride ratio was adjusted from unity to provide a system with zero net charge. Periodic boundary conditions were applied to minimize edge artifacts, and particle-mesh Ewald (PME) summation was used for long-range electrostatic interactions; short-range interactions were cut off at 12 Å.

3.4.3 Poisson–Boltzmann calculations

All solutions of the linearized-Poisson–Boltzmann equation were obtained using a multigrid finite-difference solver distributed with the ICE (Integrated Continuum Electrostatics) package.⁽¹¹⁴⁻¹¹⁵⁾ The atomic radii used were those optimized by Roux and co-workers for use with continuum electrostatic models;⁽¹¹⁶⁻¹¹⁷⁾ partial atomic charges were taken from the PARAM22/27 parameter set.⁽¹¹²⁻¹¹³⁾ For consistency with the all-atom molecular dynamics simulations, an internal dielectric constant of 1 was used, along with an external dielectric constant of 80. Two sets of calculations were performed, one with an ionic strength of 0.145 M, using a 2.0 Å ion exclusion (Stern) layer, and the other with 0.0 M salt. Binding free energies were computed as the difference between the bound complex and rigidly separated unbound

components; the bound and unbound states were positioned identically on the finite-difference grid so as to cancel artifactual grid energy. Focusing boundary conditions were used: an initial calculation was done with the largest dimension of the system occupying 23% of one side of the grid, and Debye-Hückel boundary conditions were used; a second calculation was then done with the system occupying 92% of the grid, and using the potentials from the previous calculation at the boundary. For individual components, but not for net binding energies, an additional overfocused calculation was done with the molecule occupying 184% of the grid (centered on the component of interest). Net binding energies were computed with a 257^3 unit cubic grid, and individual components with a 129^3 unit grid; due to the overfocusing used for components, the grid spacing was identical in both of these cases.

3.4.4 Generalized-Born calculations

Generalized-Born calculations were done using the GBSW module (105) of a version of the CHARMM computer program modified to output the effective Born radii. As for the PB calculations, the atomic radii used were those optimized by Roux and co-workers.(116-117) Binding was considered as the difference between the bound state and a state in which one component was rigidly translated 500 Å. The scaling coefficients were set to standard values, of $a_0 = 1.2045$ and $a_1 = 0.1866$, the molecular surface was used, and a smoothing length of 0.2 Å was applied; no cutoffs were used. GBMV calculations (Appendix A) were also performed using CHARMM,³⁴ using Lebedev integration with a grid of 38.

3.4.5 Component analysis

PB-based component analysis was done with the ICE software package, using a standard approach.(37, 114) Each amino acid was divided into three groups: side chain, amino (including CR and attached hydrogens), and carbonyl. For each group, the LPB equation was solved (in both the bound and unbound states) for a system in which only that group is charged; the difference in the potentials from this pair of calculations was used to compute the self-desolvation for the group and all interactions (direct and indirect) involving the group. GB-based component analysis was done with the same partitioning of groups. Effective Born radii for each atom in the bound and unbound states were computed with the CHARMM computer program as outlined above. These radii were then used within a separate, locally written program to compute

the pairwise energies between all groups. As every term is a difference between the bound and unbound states, intramolecular terms (desolvation and indirect interactions) include only the polarization term of Equation 3.14, while intermolecular (direct) interactions additionally include the Coulombic term for the bound state.

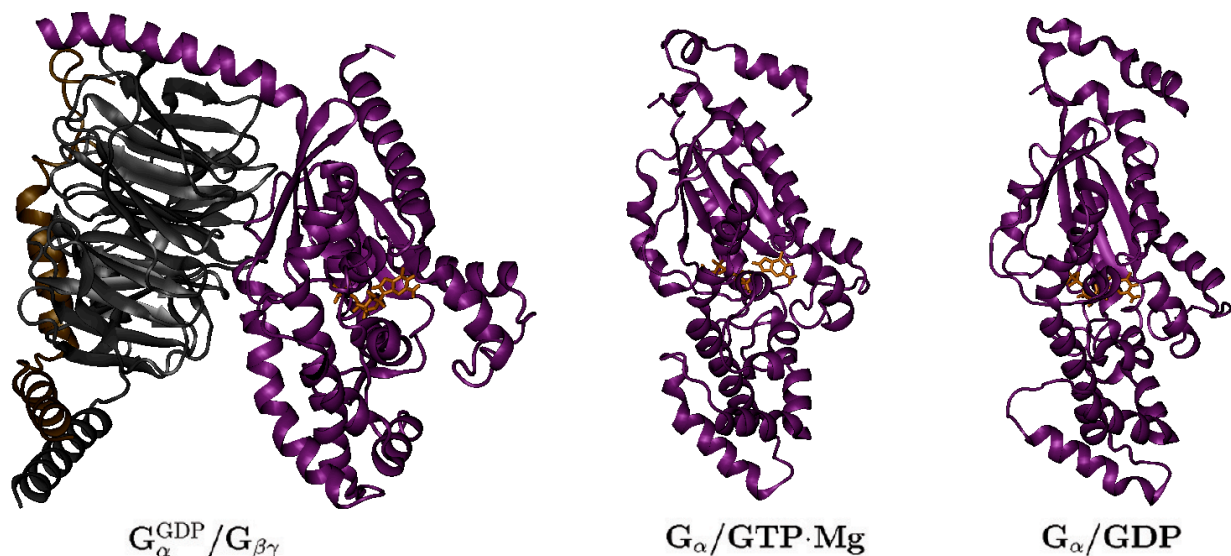


Figure 3.1 Overview of complex structures. For each system studied, a representative structure is displayed, with G_{α} in purple cartoon, G_{β} in gray cartoon, G_{γ} in bronze cartoon, and the nucleotide (GTP or GDP) in tan licorice: (left) the G-protein heterotrimer, $G_{\alpha}^{\text{GDP}}\cdot G_{\beta\gamma}$ (PDB ID 1gp2); (center) the activated G_{α} monomer, $G_{\alpha}^{\text{GTP}\cdot\text{Mg}}$ (PDB ID 1gia); (right) the G_{α} monomer in the inactive state, G_{α}^{GDP} (PDB ID 1gdd). Figures generated with VMD.78(118)

3.5 Results

The heterotrimeric G-proteins can exist in both a bound complex or an unbound state; additionally, the α -subunit may be bound to guanosine-5'-di- or triphosphate. In the unbound state, G_{α} has been structurally characterized bound to both GDP and GTP (or more precisely, a nonhydrolyzable GTP analogue), while, in the bound state, only the GDP-bound state is experimentally accessible—nucleotide exchange is coupled with trimer dissociation as a fundamental step in signaling. We have thus begun our analysis of this system with three states: isolated G_{α} bound to GDP; isolated G_{α} bound to GTP; and G_{α} bound to GDP, in complex with $G_{\beta\gamma}$.

In each of these systems, we have considered a single binding equilibrium; for the unbound α -subunits this is nucleotide association, while for the trimeric complex it is the association of $G_{\beta\gamma}$ with G_{α}^{GDP} . Each of these systems has distinct features at the binding

interface, and thus, the set of three presents a range of interaction chemistries. In particular, $G_{\alpha.GDP} - G_{\beta\gamma}$ binding is characteristic of the interactions of large proteins, with a large amount of surface area buried on binding, and a diverse range of side chain–side chain, side chain–backbone, and backbone–backbone contacts. Nucleotide binding, on the other hand, involves the near complete burial of a small organic molecule in a deeper binding pocket. The nucleotides are both highly polar molecules carrying a net negative charge ($-3e$ for GDP and $-4e$ for GTP). GTP, however, is associated in the binding site with a magnesium ion. In solution, the preferred state of nucleotide triphosphates is in complex with magnesium,⁽¹¹⁹⁾ and thus, considering the ligand as a GTP.Mg complex with net charge $-2e$ is a reasonable approximation. These structures are displayed in Figure 3.1.

3.5.1 Electrostatic binding free energies computed with generalized-Born and with linearized Poisson–Boltzmann are well-correlated

Molecular dynamic simulations of each system were performed for 100 ns in explicit solvent with periodic boundary conditions. Analysis of these trajectories indicated that they were all stable and well-behaved (see the Supporting Information). 101 evenly spaced snapshots were then extracted for the computation of binding free energies with both the generalized-Born and Poisson–Boltzmann methods. These data are shown in Figure 3.2 and Table 3.1. There is quite strong correlation between GB- and PB-computed electrostatic binding free energies in all cases, with R^2 values of greater than 80%; these correlations persist over a range of over 100 kcal/mol in all three systems. However, there are important differences as well.

3.5.2 Generalized-Born Overestimates Net Electrostatic Binding Free Energies

While for some snapshots the GB-computed binding free energy is lower than that computed with PB, overall there is a strong trend for lower binding free energies with PB. The ensemble-averaged binding free energy for the trimer is 2 kcal/mol lower when calculated with PB as compared to the same value computed with GB; the nucleotides show an even greater tendency for GB to give higher values, with a 10 kcal/mol difference for GTP·Mg binding to GR and a 22 kcal/mol difference for GDP. While these differences are within the standard deviations of the ensembles (see Table 3.1), they are not within the standard errors of the mean.

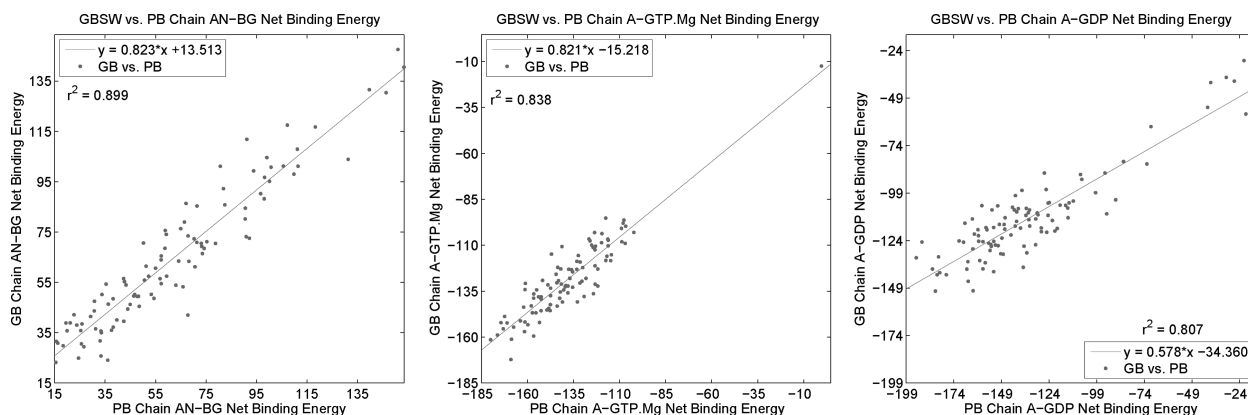


Figure 3.2 Variation in single-snapshot net binding energies computed by GB and PB. Rigid-body electrostatic binding free energies (in kcal/mol) computed for each snapshot by both GB (y -axis) and PB (x -axis) are shown for all three systems studied: (left) $G_{\alpha.GDP}.G_{\beta\gamma}$; (center) $G_{\alpha.GTP.Mg}$; (right) $G_{\alpha.GDP}$. Note that the single outlier of high energy for $G_{\alpha.GTP.Mg}$ corresponds to the first frame of the simulation; when this point is excluded, R^2 is 0.761, and the best-fit line is $y = 0.808x - 17.078$.

Table 3.1: Net electrostatics binding free energies.^a

System	mean		Std. dev		R^2	rms
	PB	GB	PB	GB		
$G_{\alpha}^{GDP} / G_{\beta\gamma}$	62.16	64.70	32.26	28.01	0.90	10.82
$G_{\alpha} / GTP.Mg$	-138.38	-128.81	22.98	20.61	0.84 ^b	13.30
G_{α} / GDP	-135.33	-112.63	36.11	23.25	0.81	29.13

^aAll energies in kcal/mol; all values are computed over a set of 101 evenly spaced snapshots from a 100 ns MD trajectory. ^b R^2 is 0.76 when the first snap shot of the trajectory is excluded.

3.5.3 Generalized-Born underestimates differences in net electrostatic binding free energies between configurations

Considering the best-fit lines relating the two values makes it clear that the reason for these discrepancies is not simply a difference of reference state. In all cases, the least-squares fit describing the variation of GB-computed binding energies to those computed with PB has a slope less than unity—slightly above 0.8 for association of the trimer and for the binding of GTP.Mg to G_{α} and below 0.6 for GDP— G_{α} binding. A slope of below unity indicates that GB will tend to underestimate the magnitude of free-energy differences between configurations, as compared to PB—both highly favorable (negative) and unfavorable (positive) values will be less so with GB. While this general trend holds in all three cases, the offset from zero (intercept) is very different in all cases. For the association of the protein components of the trimer, where the net electrostatic binding energy is always positive, the intercept is also significantly positive (+15 kcal/mol); it is this positive offset that makes the mean GB result greater than that from PB,

despite a slope of correlation below $y = x$. For the binding of the nucleotides, however, the intercept is negative (-15 kcal/mol for GTP·Mg and -34 kcal/mol for GDP). Despite this, the highly negative values of the binding free energies make the less than unit slope dominate the difference in the means. Although there are significant variations between the equations of the best-fit lines in the different systems, fitting the data from all three components gives a strong correlation ($R^2 = 0.98$) with a least-squares fit having a slope of 0.91 and intercept of 5.3 (see Appendix A). Thus, gross differences between systems seem to be more closely captured by the two models than are the details of energetic variation within a system.

3.5.4 Energetic components show highly variable correlations between GB and LPB models

As discussed in the Theory section, one of the great benefits of linear-response models such as generalized-Born and linearized Poisson–Boltzmann is that computed energies may be rigorously decomposed into contributions from individual groups and pairs of groups. In order to better understand the observed differences in binding free energies computed with the two models, we performed such a decomposition with both approaches. Each amino acid was partitioned into three groups: side chain, backbone amino, and backbone carbonyl. For each group, the difference in the self-energy in the bound and unbound states (due to differences interactions with solvent) is termed the desolvation energy; for each pair of groups within a molecule, the difference in solvent screening of the interaction (in the bound and unbound states) is termed the indirect interaction; and for each pair of groups on opposite binding partners, the solvent-screened Coulombic interaction is termed the direct interaction. Finally, the sum of all terms involving a single group (desolvation and all interactions, both direct and indirect) is termed the mutation energy; the mutation energy is the energetic difference in binding free energy between the natural system and a hypothetical system with that group (and that group alone) replaced with a hydrophobic isostere. Figure 3.3 and Table 3.2 summarize these data, excluding components for which both GB and PB are within ± 1.0 kcal/mol of zero; these consist primarily of components not close to the binding interface.

Table 3.2. Group-Based Electrostatic Binding Energy Component.^a

system	N	Mean		Std. dev.		R^2	rms
		PB	GB	PB	GB		

Mutation Energy							
$G_{\alpha.GDP.G_{\beta\gamma}}$	9485	-1.70	-1.76	5.81	5.57	0.91	1.75
$G_{\alpha.GTP.Mg}$	7328	-6.30	-6.22	18.18	17.35	0.93	4.71
$G_{\alpha.GDP}$	9621	-4.21	-3.71	14.85	11.30	0.84	6.31
Desolvation Penalties							
$G_{\alpha.GDP.G_{\beta\gamma}}$	5505	4.67	4.41	5.17	5.08	0.79	2.43
$G_{\alpha.GTP.Mg}$	1341	5.75	3.43	5.23	4.14	0.83	3.22
$G_{\alpha.GDP}$	1439	5.33	3.69	5.34	4.51	0.64	3.60
Indirect Interactions							
$G_{\alpha.GDP.G_{\beta\gamma}}$	5294	-1.23	-1.09	3.25	1.87	0.48	2.38
$G_{\alpha.GTP.Mg}$	5710	0.68	0.84	5.02	4.44	0.90	1.65
$G_{\alpha.GDP}$	4765	0.31	1.08	4.59	5.36	0.51	3.89
Direct Interactions							
$G_{\alpha.GDP.G_{\beta\gamma}}$	9490	-3.87	-3.84	8.67	8.60	0.98	1.26
$G_{\alpha.GTP.Mg}$	10342	-5.74	-5.38	18.99	17.07	0.96	4.30
$G_{\alpha.GDP}$	10812	-4.68	-4.36	16.85	14.72	0.94	4.41
^a All energies are in kcal/mol. Values are computed over the set of components from all 101 snapshots with either a GB- or PB-computed energy of greater than 1.00 kcal/mol in absolute value; N is the size of this set.							

3.5.5 Direct, solvent-screened Coulombic interactions show bimodal correlations

In considering the mutation energies, a very interesting result appears. The great majority of points show strong correlation in the values computed by each method, with the data largely falling along a line of near unit slope. However, in all systems, there appears to be a subpopulation with a decidedly lower slope. This is most apparent for the nucleotide-binding systems but is also seen for a few groups in trimer association.

The mutation energy includes both indirect (intramolecular) and direct (intermolecular) terms, as well as self-desolvation penalties, and it is informative to consider how each of these vary when computed in the two models. Individual group desolvation energies computed with the two models are only loosely correlated. On average, generalized-Born underestimates individual desolvation energies; this is by a small amount for the components of the trimer but a larger amount for the protein components of the nucleotide-binding systems. However, visual

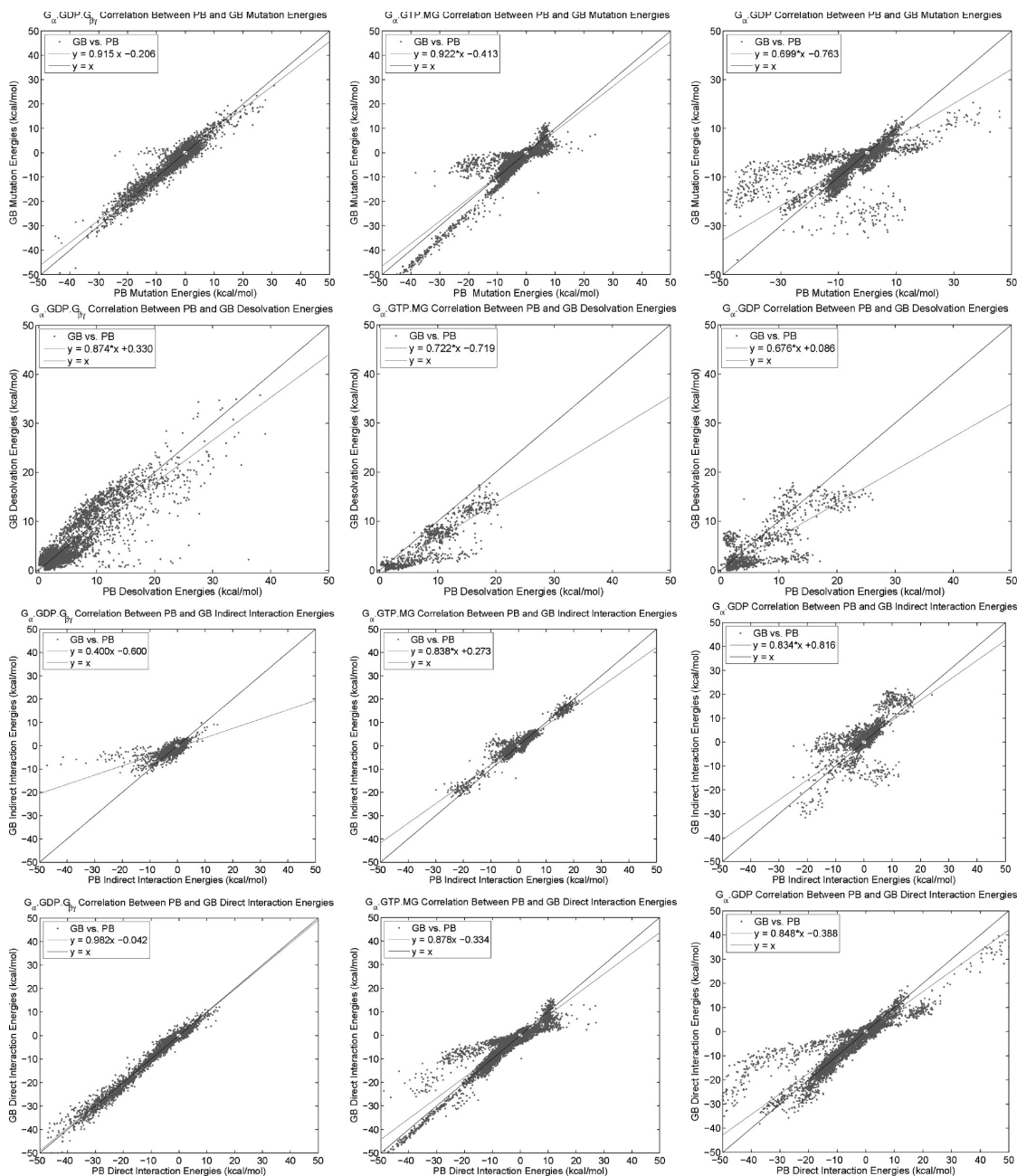


Figure 3.3 Variation in group-based energetic terms computed by GB and PB. The individual group-based energetic terms computed for every snapshot of each system are displayed. Columns correspond to the three different systems: (left) $G_{\alpha.GDP.G_{\beta}}$; (center) $G_{\alpha.GTP.MG}$; (right) $G_{\alpha.GDP}$. Rows correspond to different energetic terms: (top) mutation energies (relative to hydrophobic isostere); (second) desolvation penalties; (third) indirect (intramolecular) interactions; (bottom) direct (intermolecular) interactions. All energies are in kcal/mol, and for clarity, components with energies computed to be within ± 1.0 kcal/mol by both methods have been excluded, and the range plotted is restricted to ± 50.0 kcal/mol.

analysis of these data clearly reveals that the deviations are not randomly distributed. Indirect interactions, corresponding to the differential solvent screening of intramolecular interactions in

the bound and unbound state, show relatively weak correlations, although the degree of correlation is system-dependent. The direct interactions, on the other hand, show the greatest degree of correlation. The direct interactions also most clearly exhibit the existence of two populations, with different slopes of correlation; this seems to be the primary but not exclusive source of the patterns seen in the mutation term.

3.5.6 Inconsistencies between GB and PB are group-specific

That certain groups show qualitatively different behaviors raises the possibility that the distinct properties of particular amino acids may play a role. To consider this, the contributions of each group were further broken down by group type. Figures 3.4 and 3.5 show these data for the association of $G_{\alpha, \text{GDP}}$ and $G_{\beta\gamma}$ to form the G-protein heterotrimer.

Individual group solvation free energies (Figure 3.4) show a wide range of behaviors, with some groups showing strong correlation between the GB- and PB-computed values (10 of the 39 most significant groups have R^2 values of greater than 0.80) and others showing minimal correlation (R^2 is below 0.5 for 13 of the 39 groups with significant desolvation energies by PB). For example, essentially all glutamate and lysine points are distributed about $y = x$, albeit with some degree of variation. Glutamine, tyrosine, and arginine also display relatively well distributed linear correlations but with slopes somewhat below unity; the backbone groups (both amino and carbonyl) also tend to follow this pattern. However, some groups show much more structure in the variation between the two methods. The data for asparagine, for example, fall into distinct clusters, largely identified as particular residues; the data for Asn B88, for example, are well-correlated ($R^2 = 0.88$) and approximately follow $y = x$ (with a slight offset), while the data for Asn B119 are more weakly correlated ($R^2 = 0.40$) with a slope significantly below 1. For serine and threonine, most positions show very little correlation but in a manner that varies; Ser A16 and A206, as well as Thr A182 and B153, all have GB-computed values largely confined to about a 1 kcal/mol range near 3 kcal/mol, while their PB-computed values span a wider range (1-6 kcal/mol). In other cases, however, some degree of correlation is seen, although generally with a slope much less than 1. The most profound deviation between the GB- and PB-computed values is for the protonated histidine at B54, with a correlation coefficient of only 0.18; by Poisson-Boltzmann, the desolvation of this group ranges from about 5 to 25 kcal/mol, while, with GB, the desolvation never exceeds 5 kcal/mol.

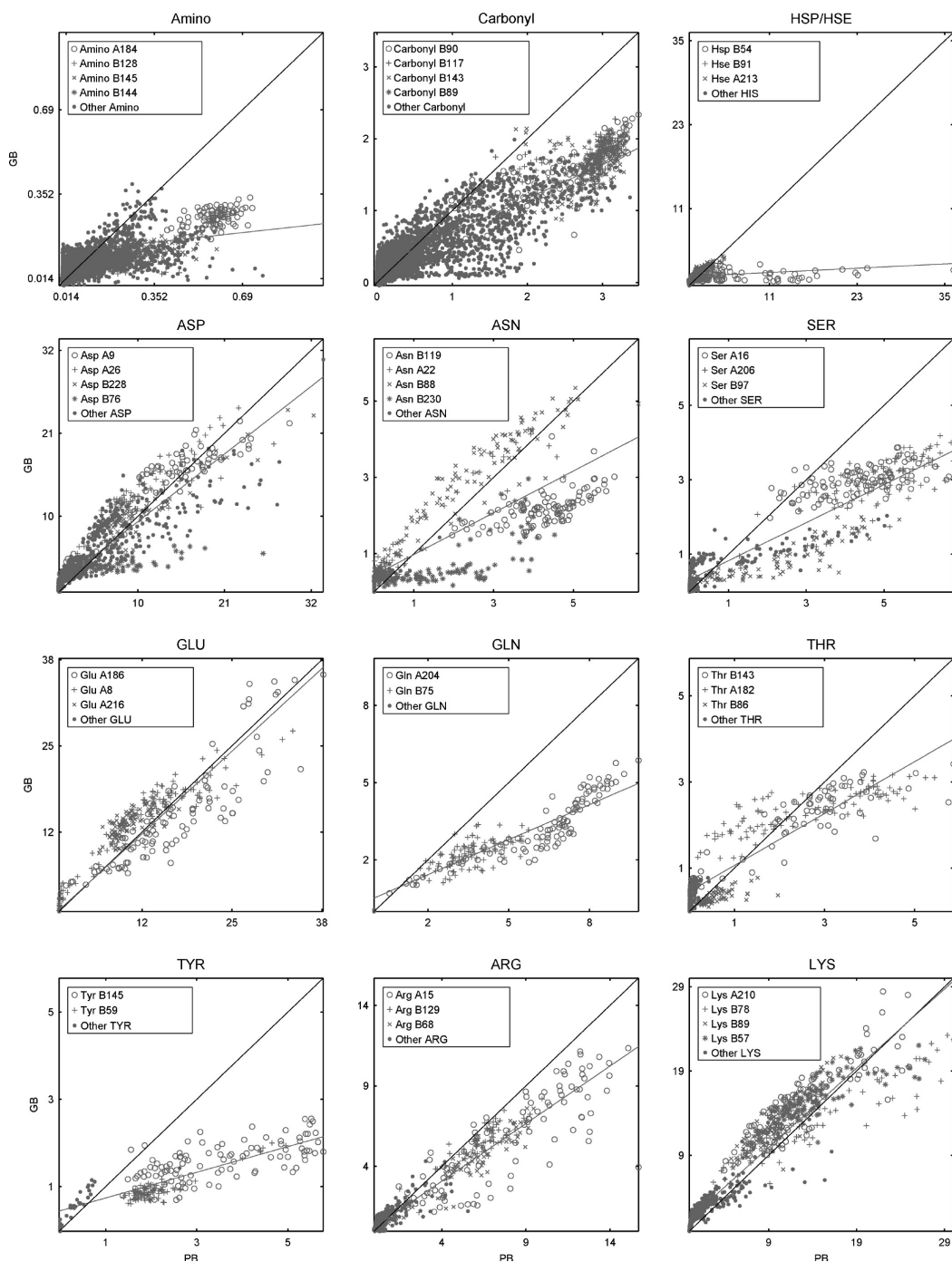


Figure 3.4. $G_{\alpha\text{-GDP}} \cdot G_{\beta\gamma}$ group desolvation penalties, by type. The data displayed in Figure 3.3 (second row, left) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no cysteines make interactions across the interface, these are also not shown.

3.5.7 Contributions of a buried salt bridge are missed by GB

As noted earlier, group mutation energies in this system show a remarkable level of correlation given the many inconsistencies in the desolvation terms, and this is true for essentially all group types (Figure 3.5); 27 of the 39 most significant groups (by PB) have correlation coefficients between the two models of greater than 0.80. There are, however, two notable exceptions to this: Asp 76 and His⁺ 54 on the β -subunit. Both of these groups are computed in the Poisson–Boltzmann model to make significantly favorable contributions in a subset of snapshots, of as much as -25 kcal/mol for Asp 76 and -15 kcal/mol for His 54, while the terms computed with GB are uniformly low in magnitude. As seen in Figure 3.6, these two groups are located proximal to one another, near—but not directly involved in—the interface with G_α . In the crystal structure, they are involved in an intramolecular salt bridge with one another; through the course of the simulation, this salt bridge breaks and then reforms. In all cases, however, both groups make additional intramolecular interactions that keep the groups localized beneath the protein surface. Under the Poisson–Boltzmann model, the desolvation penalty paid by both groups is substantial for certain snapshots, while the GB model suggests a small desolvation, uniform across all snapshots; this leads to a correlation coefficient between the two models of only 0.24 for His and 0.05 for ASP, and the best fit lines have a slope of only 0.20 and 0.04, respectively. In the PB model, the indirect interactions made between these groups also are very large in some cases (particularly when the salt bridge is formed); these correspond to the enhancement of the favorable electrostatic interactions between the two groups upon removal of solvent from the volume of the α -subunit, and the GB model is unable to capture these effects.

3.5.8 Deviations between PB and GB are larger for nucleotide binding

The differences seen between the two methods in assessing the contributions of individual groups to the association of the G_α and $G_{\beta\gamma}$ subunits are generally moderate, particularly due to the cancellation of error in the mutation term. However, this is much less so in the case of nucleotide binding. Figures 3.7 and 3.8 show these data for the association of G_α with GTP.Mg. In this system, self-desolvation energies computed with GB are greatly underestimated (with respect to the values from PB) for the majority of groups with significant desolvation. This is seen for serine and threonine, as well as for cysteine, asparagine, although

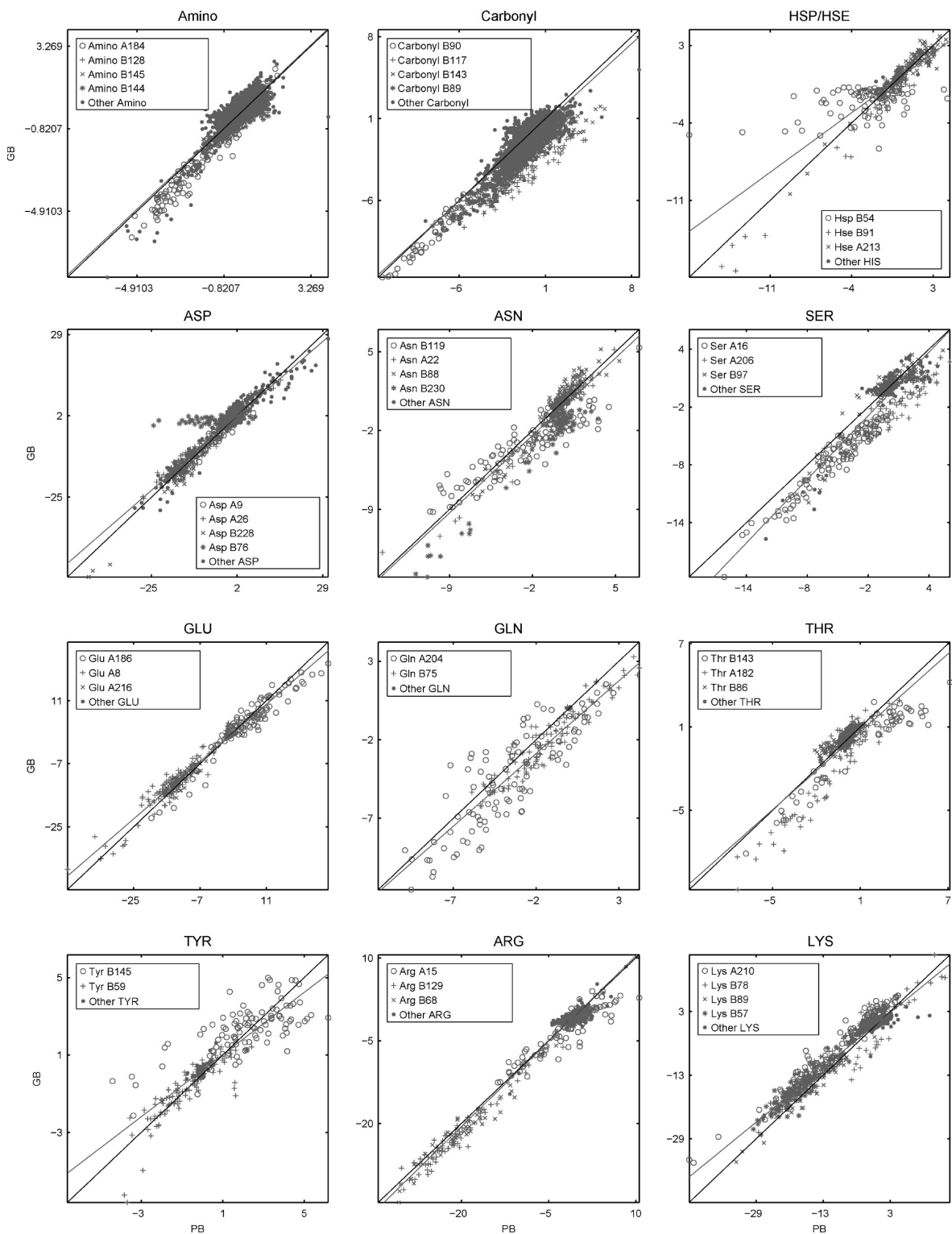


Figure 3.5. $G_{\alpha, \text{GDP}} \cdot G_{\beta \gamma}$ group mutation energies, by type. The data displayed in Figure 3.3 (first row, left) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no cysteines make interactions across the interface, these are also not shown.

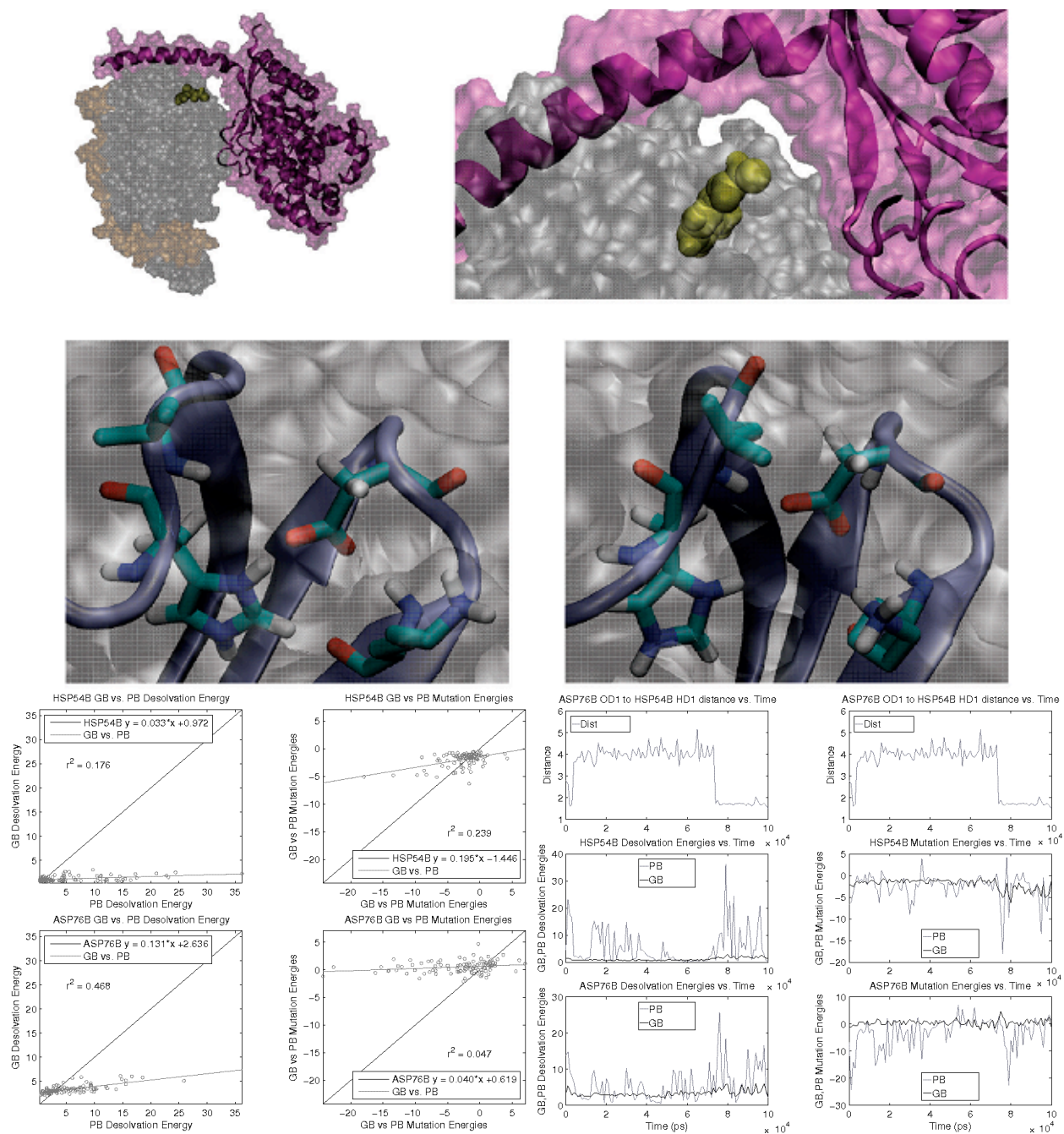


Figure 3.6. Structure of His⁵⁴ and Asp76 of G β in the G α .GDP.G $\beta\gamma$ complex. Top: This pair of residues is located proximal to the G α binding interface (left) but is oriented so as not to make direct interfacial contacts, nor to be solvent exposed (right). Center: In some snapshots, the two groups make a hydrogen-bonded salt bridge, with a representative structure shown on the left; in others, the aspartate shifts to interact with the backbone amino group of Leu55. Bottom: The energetics of each component are shown, with the left-most panels displaying the correlation of GB- and PB-computed desolvation and mutation terms for each residue (His⁵⁴ top, Asp76 bottom) and the right-most panels showing the variation of these terms during the course of the simulation; the top graph of the left-most panels shows a metric of salt bridge formation. Structural figures generated with VMD.(118)

no terms are particularly large for the latter three. Similar behavior is seen for the backbone groups. For individual groups, even when there is significant correlation, the best-fit line has a very low slope; the GB- and PB-computed results for Ser 151 are correlated with an R^2 value of 0.50, but the best-fit line has a slope of 0.15, and for the backbone NH of position 45 the R^2 value is 0.36, but the best fit slope is 0.20 (correlation coefficients and best-fit lines for all significant groups are given in the Supporting Information). For both aspartate and arginine, reasonable agreement between the two methods is seen, although with significant variances some positions are highly correlated with near unit slope (for example, Asp 272 with $R^2 = 0.96$, slope 0.84, and intercept -0.1), while others are less so (such as Asp 200, with $R^2 = 0.30$, slope 0.38, and intercept 3.3). For lysine, while a (weak) overall positive correlation is seen when all positions are considered, when individual positions are considered, the GB-computed values span a very small range as compared to the PB-computed values; with Lys 51, for example, the two methods give a correlation coefficient of only 0.27, and the best-fit line has a slope of 0.24 and an intercept of 8.54. Finally, the two methods give results with very little correlation for the single glutamate that is desolvated; the data for Glu 43 have an R^2 value of 0.27.

Again, the mutation terms agree much better when computed in the two different models. For the backbone groups, as well as serine, threonine, aspartate, arginine, and lysine, the GB-computed mutation energies are linearly correlated to those computed with PB with quite a tight distribution. For example, while the desolvation energies computed with the two methods for Ser 47 were completely uncorrelated ($R^2 = 0.04$), the mutation terms are correlated with an R^2 value of 0.98; the data for the backbone of position 45, discussed earlier, shifts from a weak correlation with slope 0.2 for desolvation to a strong correlation ($R^2 = 0.79$) with a near perfect $y = x$ correlation (best-fit line, $y = 1.02x - 0.09$). However, while in the case of the α - β γ subunit association was largely due to a cancellation of errors in the computation of desolvation and interaction terms, in this case, the primary effect seems to be the dominance of the interaction terms, which are highly correlated in the two models. This can be seen clearly in the magnitude of the terms; for example, the (poorly correlated) desolvation energies for serine range from 0 to 4.5 kcal/mol, while the (highly correlated) mutation energies span a range of -40 to +10 kcal/mol. As seen in Figure 3.9, Ser 47 makes a hydrogen bond with the α phosphate of GTP; in the bound state, this position is completely buried from solvent. In a buried site, the direct

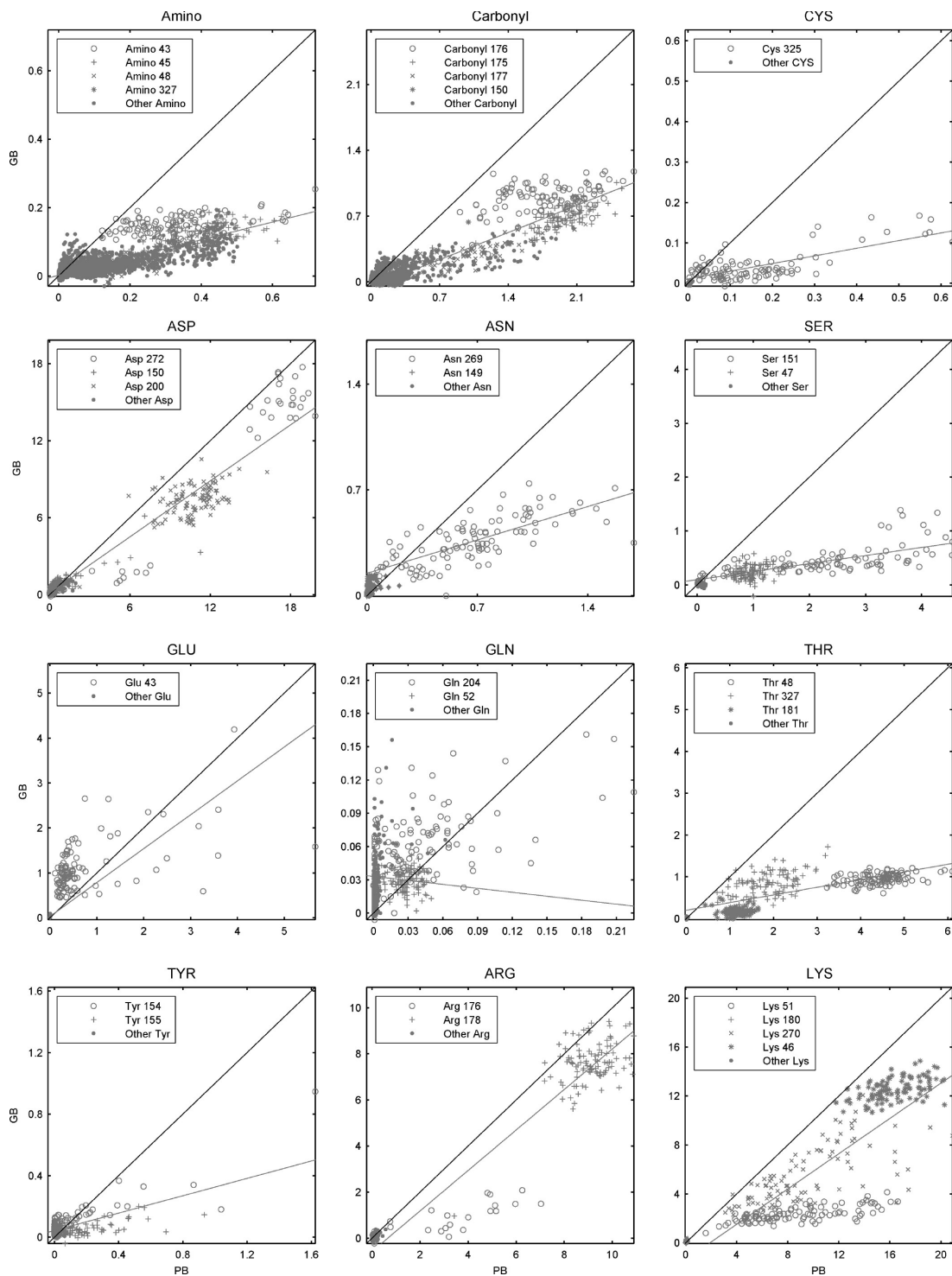


Figure 3.7. $G_{a,GTP,Mg}$ group desolvation penalties, by type. The data displayed in Figure 3.3 (second row, center) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no histidines make interactions with the nucleotide, these are also not shown.

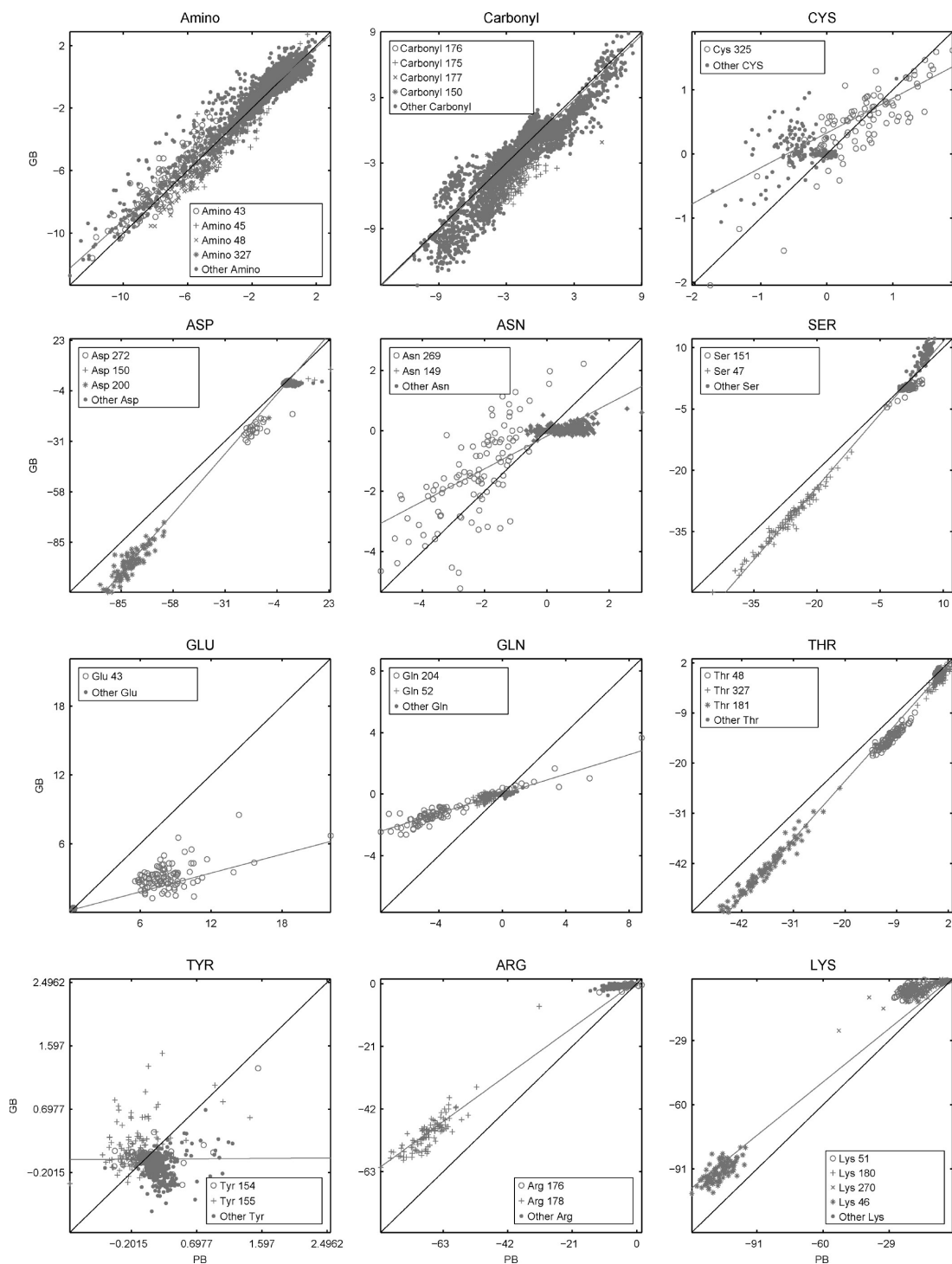


Figure 3.8. $G_{\alpha.GTP.Mg}$ group mutation energies, by type. The data displayed in Figure 3.3 (first row, center) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no histidines make interactions with the nucleotide, these are also not shown.

interaction energies will approach the limit of Coulomb's law in a uniform dielectric constant equal to that of the protein interior (in this case, 1.0). Both GB and PB approach this limit in this case, and thus agree well. These terms are large in magnitude because of the highly charged nature of the ligand, and the low internal dielectric constant. Adding to this, the desolvation penalties of many of the groups lining the binding pocket are rather low, due to significant burial in the unbound state.

3.5.9 Moderate-range, action-at-a-distance interactions are poorly captured by GB

In other cases, however, greater deviations between the GB- and PB-computed mutation energies are seen. Gln 204 pays essentially zero desolvation, and thus, the mutation energies are again dominated by interactions. However, the correlation of the values computed with the two models, while linear, has a slope much lower than 1; similar results are seen for certain asparagine, arginine, and lysine residues. As can be seen in Figure 3.9, Gln 204 does not make direct contact with the GTP ligand but does make moderate-range, action-at-a-distance electrostatic interactions with the triphosphate. The GB model overestimates the solvent screening of this interaction, as compared to the PB-computed values. The results for Glu 43, which showed no correlation in desolvation, remain essentially uncorrelated in the mutation term. The desolvation of this group is quite small for a charged group, and the interactions are unfavorable. This group is near the phosphate of the nucleotide but not involved in any direct interaction; the unfavorable action-at-a-distance effects of this group are not well-captured by GB. Lys 51, noted above as being weakly correlated with a slope much below unity for desolvation, behaves similarly in mutation energies, with only a slightly improved correlation coefficient of 0.50; this position is proximal to the phosphate tail of the nucleotide but does not make direct hydrogen-bonding interactions.

3.5.10 Structurally similar interactions are reproduced between systems with both methods

Structurally and chemically, the binding of GDP to G_{α} is very similar to that of GTP.Mg, and thus, it may be expected that the behavior of the two models in these systems would also be similar. The data for G_{α} —GDP binding are shown in Figures 3.10 and 3.11, and while many

similarities are seen, there are notable differences. The desolvation energies of Glu 43, showing very little correlation with GTP.Mg binding, are strongly correlated ($R^2 = 0.90$) with a reason-

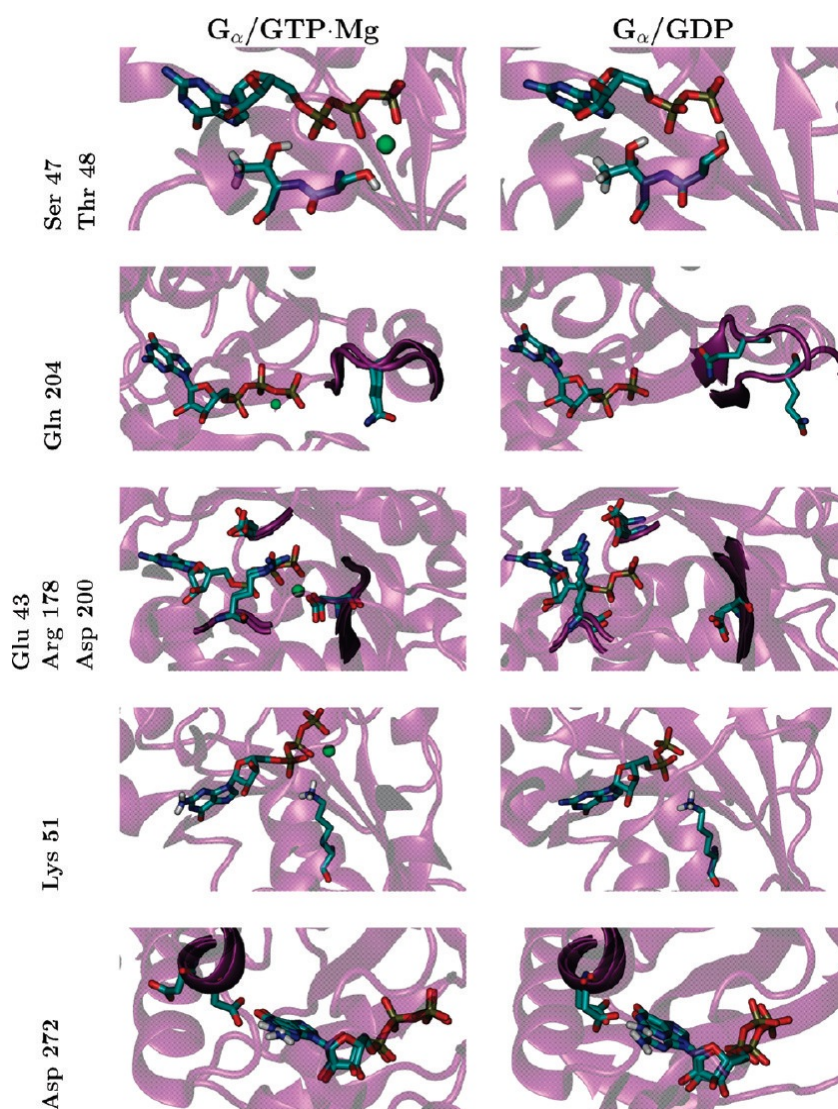


Figure 3.9. Structure of key G_{α} residues in the $G_{\alpha.GTP.Mg}$ and $G_{\alpha.GDP}$ complexes. Ser 47 makes consistent interactions with the α phosphate in both complexes; Thr 48 coordinates the magnesium ion through the oxygen of the side chain in the complex with GTP.Mg but directly donates a hydrogen bond to the β phosphate in the GDP complex. Gln 204 makes consistent action-at-a-distance electrostatic interactions with GTP.Mg but is structurally variable in the GDP-bound state. Arg 178 shifts from making a bidentate interaction with the β and γ phosphates of GTP to forming an intramolecular salt bridge with Glu 43 when GDP is bound; Asp 200 coordinates the magnesium of GTP.Mg but makes only repulsive, action-at-a-distance interactions with GDP. Lys 51 makes no direct nucleotide contacts but makes favorable action-at-a-distance interactions with both GTP and GDP. While in the GTP-bound state, Asp 272 occupies two structural clusters (one interacting with the nucleotide base and the other not); in the complex with GDP, it makes consistent hydrogen-bonded interactions. Figures generated with VMD.(118)

able slope (0.68) with GDP; while a similar correlation remains in the mutation terms ($R^2 = 0.85$), the slope is significantly reduced (0.36). In this state, Glu 43 makes a hydrogen-bonded salt bridge with Arg 178; this interaction covers the nucleotide diphosphate and thus brings the aspartate closer, making correspondingly stronger repulsions. These more intimate interactions seem to be better captured by GB, although there are still significant discrepancies.

The desolvation energies of lysines and arginines computed by the two methods appear, on visual inspection, to be less correlated with GDP binding. However, more careful analysis shows that the dominant observation is simply that the various groups have more similar mean desolvation penalties; there is little correlation for the data of a single position in either case, but with GTP.Mg binding, the wider range of contributions from different positions results in a better apparent correlation.

3.5.11 Quantification of energetic differences between systems can be very model dependent

When the interactions made in the two systems are similar, comparisons between the model generally give consistent results within an individual system, as seen in the case of Ser 47. Thr 48 also behaves similarly with both GTP.Mg and GDP binding, and is worthy of note as the details of the interaction change in the two systems. While in the GTP·Mg-bound state, the side chain oxygen coordinates the metal ion, in the GDP-bound form, the side chain hydroxyl donates a hydrogen bond directly to the terminal (β) phosphate (Figure 3.9). These interactions differ in important ways, but both are direct, short-range interactions, where the correlation of the two models seems to be greatest. However, even when energetic differences between structures of the same system are captured similarly, there can be striking differences between systems; the difference in the ensemble-averaged mutation energy of Ser 47 for GTP.Mg binding versus GDP binding is computed to be +4.3 kcal/mol by PB (favoring GTP.Mg), while the same difference computed by GB is +13.2 kcal/mol. For Thr 48, on the other hand, the same difference is -0.6 kcal/mol by PB and -1.8 kcal/mol by GB, a much closer agreement. In the correlations of GB- and PB-computed mutation energies, there are additional differences between the two nucleotide-binding systems. Arg 178 and Asp 200 both make vastly reduced contributions to GDP binding than to GTP.Mg. When binding GTP.Mg, Arg 178 makes a bidentate interaction

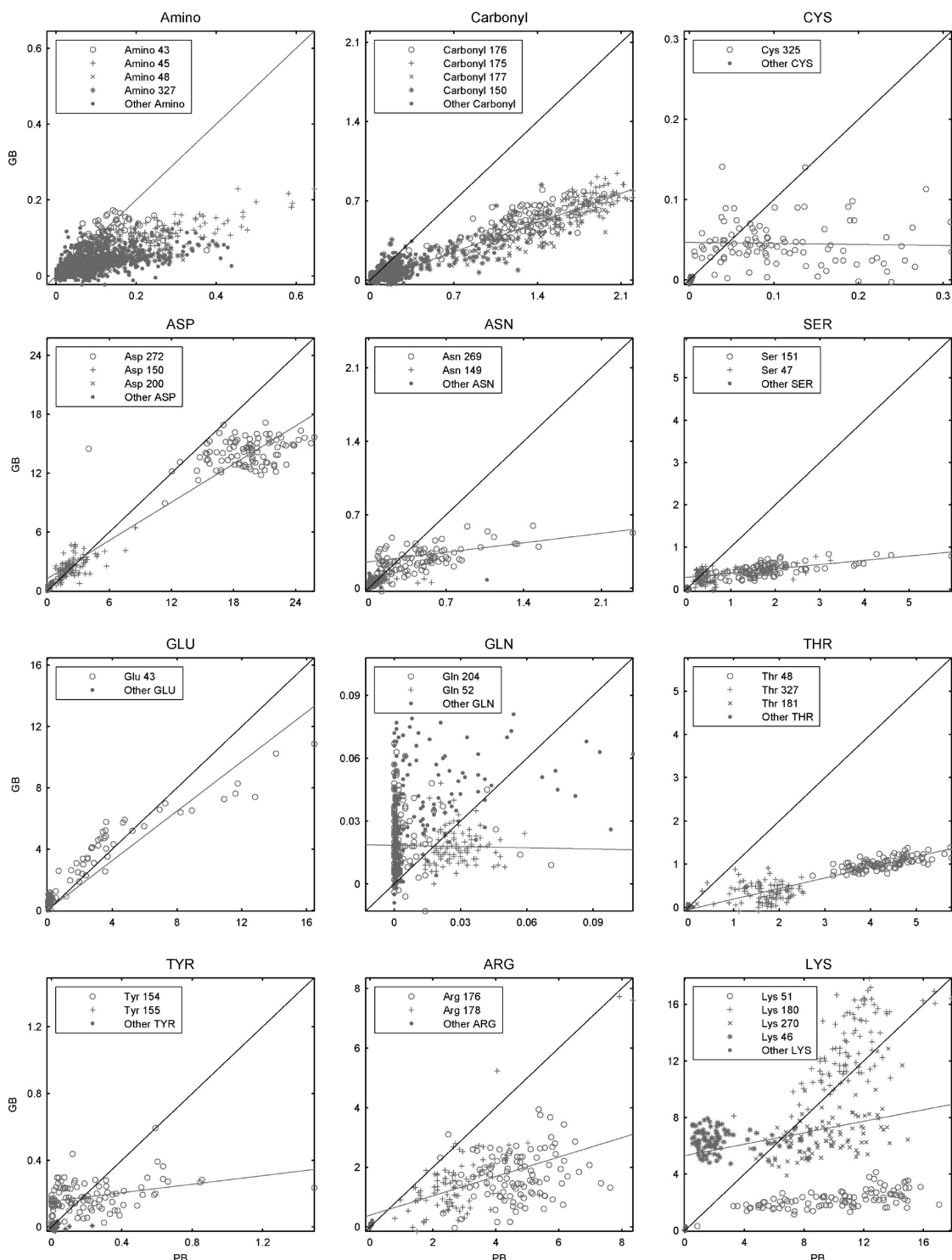


Figure 3.10. $G_{a,GDP}$ group desolvation penalties, by type. The data displayed in Figure 3.3 (second row, right) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no histidines make interactions with the nucleotide, these are also not shown.

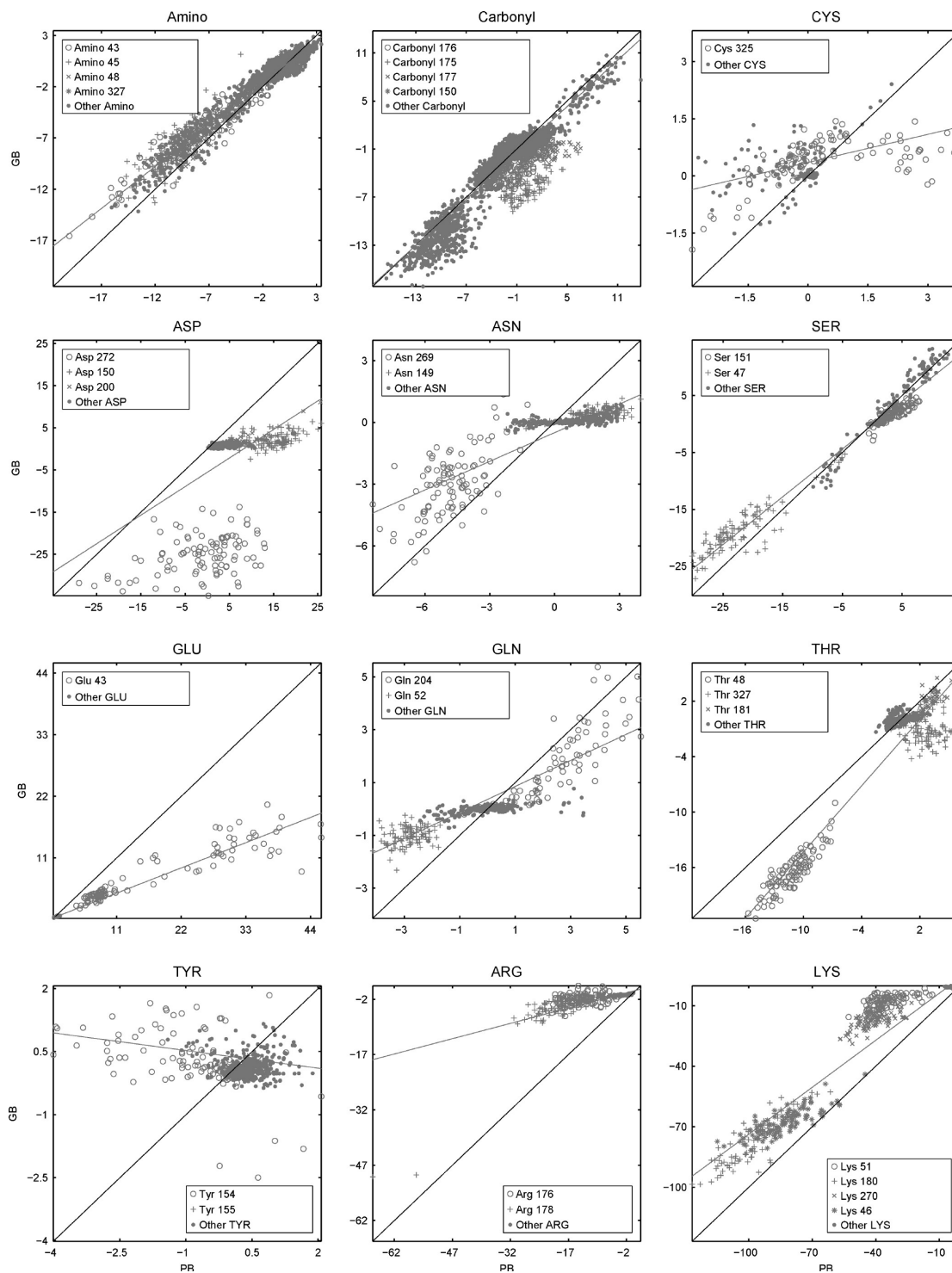


Figure 3.11. $G_{\alpha,GDP}$ group mutation energies, by type. The data displayed in Figure 3.3 (first row, right) are shown, further broken down by group type. In each case, individual residues making significant energetic contributions are uniquely marked. Nonpolar groups (aliphatic and aromatic amino-acid side chains) do not make significant electrostatic contributions and thus are not shown; as no histidines make interactions with the nucleotide, these are also not shown.

with the β and γ phosphates and Asp 200 coordinates the magnesium, but with the loss of both the γ phosphate and the associated Mg^{2+} , these interactions are lost (Figure 3.9). While Arg 178 becomes involved in an intramolecular interaction with Glu 43 (discussed above), Asp 200 becomes partially solvent exposed. As it was these groups that dominated the agreement between GB- and PB-computed mutation energies, the correlations drop greatly. Again, these unfavorable action-at-a-distance interactions are not well captured. However, while the differences within a structure do not agree well, the gross differences between the states are in general agreement. PB gives an ensemble-averaged mutation energy for GTP.Mg binding of -79.5 kcal/mol for Asp 200 and -67.5 kcal/mol for Arg 178, and corresponding terms for GDP binding of +5.6 and -19.5 kcal/mol; with GB, these values are -95.6 and +1.7 kcal/mol for Asp 200 and -50.0 and -4.8 kcal/mol for Arg 178 (complete data in Appendix A). Thus, the differences agree to within about 13% for Asp 200 and 6% for Arg 178. Asp 272 shows similar behavior but in the opposing direction and somewhat less pronounced. This group makes direct hydrogen-bonding interactions with the nucleotide base, which persist throughout the simulation with GDP bound, but which is broken for some of the GTP.Mg-bound simulation. The rough differences in desolvation energy between these two states are captured by GB (ensemble averages of 4.3 kcal/mol for GTP binding and 14.1 kcal/mol for GDP binding) in a somewhat similar fashion as by PB (5.2 and 19.3 kcal/mol, respectively), but while PB shows decided variations within each cluster, GB does not. Additionally, while average mutation terms agree reasonably well for GTP binding (-3.5 kcal/mol with PB, -6.4 kcal/mol with GB), there are dramatic differences with GDP binding—a weakly favorable contribution of -1.2 kcal/mol is computed with PB, while GB predicts a strongly favorable contribution of -25.0 kcal/mol. Many of the points for all three cases fall into the cluster of the majority of other arginines and aspartates lying along a line of very low slope. This low-slope behavior has been discussed previously in the text.

3.5.12 Generalized-Born is an effective filter for the selection of important groups

While the results discussed above indicate that there are clearly concerns with using GB as a substitute for PB in a detailed component-based decomposition, another application is to use GB as a screen for those components expected to be most significant. These terms could then be considered in detail with PB-based calculations, while eliminating the need to perform these expensive calculations on those groups that do not make significant interactions. To test the

utility of this, the ensemble-averaged mutation energy as computed by each method was compared for each group (Figure 3.12, top). The mean energies are highly correlated in all systems, although individual residues show differences of as high as 10 kcal/mol or more.

As a direct measure of the utility of the GB-computed energies as a screen, the enrichment in the most significant components (in mutation energy) computed by PB was measured, using a GB-ranked list of components (Figure 3.12, middle and bottom). GB performs remarkably well; when a cutoff of 1.36 kcal/mol is used to determine significance (corresponding to a 10-fold difference in affinity relative to the hydrophobic isostere), 100% of the set is found within the top third of GB-ranked components in all systems. Perhaps more remarkably, 60-70% of the set can be found with very little wasted effort; among these top GB-ranked components, no more than three components (out of between 30 and 60) are not in fact significant.

3.5.13 Salt effects are well correlated between LPB and GB models, when significant

The data presented above was all from computations done in the context of 0.0 M ionic strength; in this case, the Poisson–Boltzmann equation reduces to the Poisson equation. To evaluate the effect of this, net binding free energies for all snapshots were additionally computed with a 0.145 M concentration of monovalent ions (NaCl) using both the GB and linearized-PB models; the data are summarized in Table 3.3 and Figure 3.13. Inclusion of salt has a relatively small effect on the net binding free energies; in all three systems, the correlation coefficient between the energies computed with and without salt is greater than 99.9%. For nucleotide binding, the rms difference is roughly 0.4 kcal/mol, and the maximum absolute difference is 1.0 kcal/mol or less. In these cases, where the salt effects are relatively small, the salt effects computed by the two methods are only weakly correlated. The effect of salt in the computed protein-protein binding free energy for trimer formation is much more significant; the rms difference over all snapshots is almost 3 kcal/mol, and individual differences as great as 4.5 kcal/mol are seen. However, in this case, the effects are very similar when computed with the LPB and GB models; the differences are correlated with an R^2 value of 0.88, an rms difference of 0.3 kcal/mol, and a maximum difference of only 0.7 kcal/mol. As salt effects are either relatively weak (nucleotide binding) or highly correlated between the two models (trimer

formation), the conclusions drawn from the analysis of the salt-free systems may be expected to transfer well to salt-containing calculations.

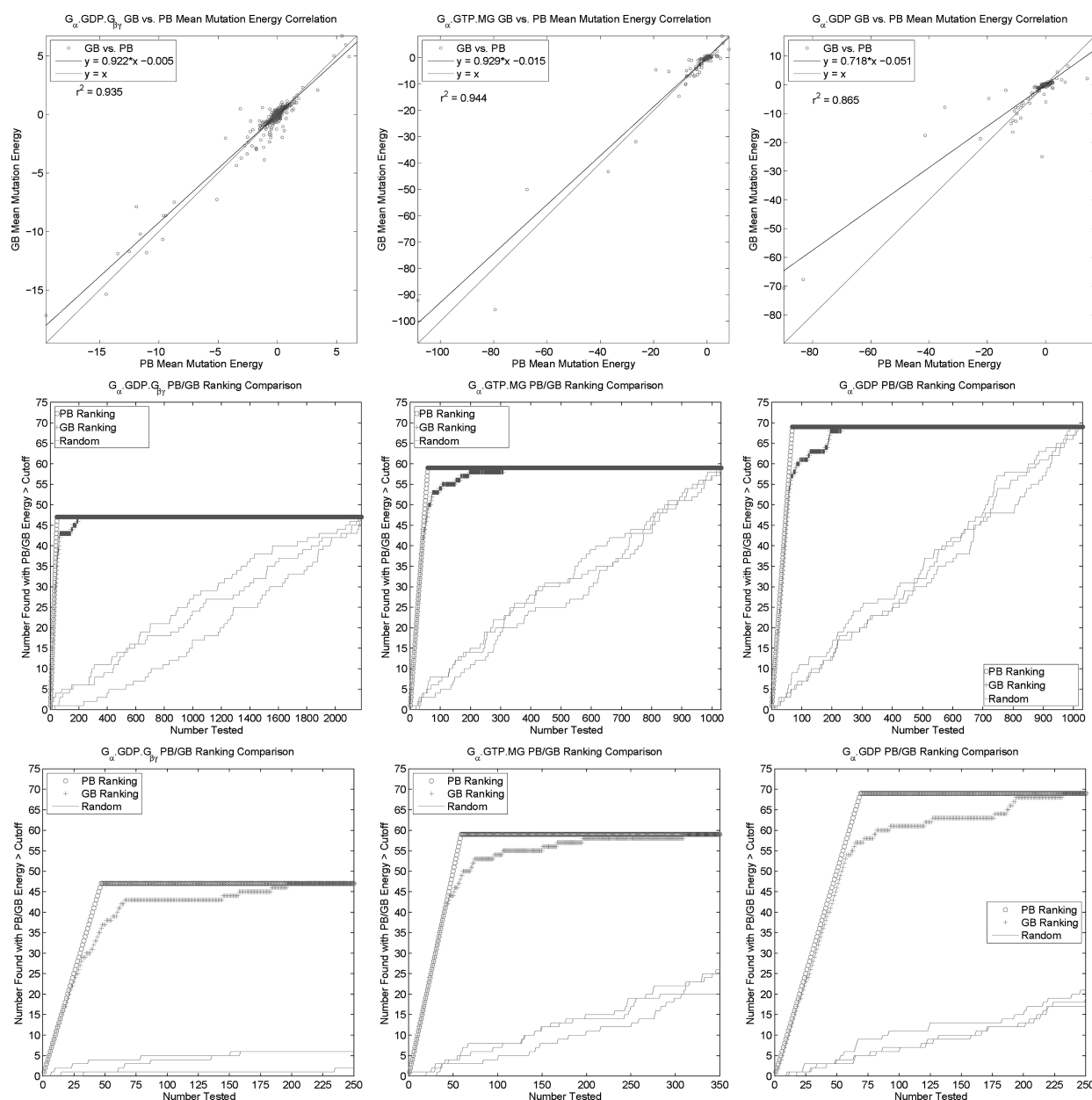


Figure 3.12. Performance of GB in selecting most significant residues by PB. The utility of GB in screening for those components that are most significant by PB is displayed in two ways for each system: (left) $G_{\alpha.GDP.G_{\beta}}$; (center) $G_{\alpha.GTP.MG}$; (right) $G_{\alpha.GDP}$. Top: The ensemble-averaged mutation energies for each component computed with GB are plotted against the same terms computed with PB. Middle and Bottom: The cumulative number of components with PB-computed mutation energies of greater than 1.36 kcal/mol in magnitude are plotted against the number of components tested, with components ranked by PB, GB, or randomly. The bottom panels show the same data as the middle panels, with the x -axis scaled for increased resolution.

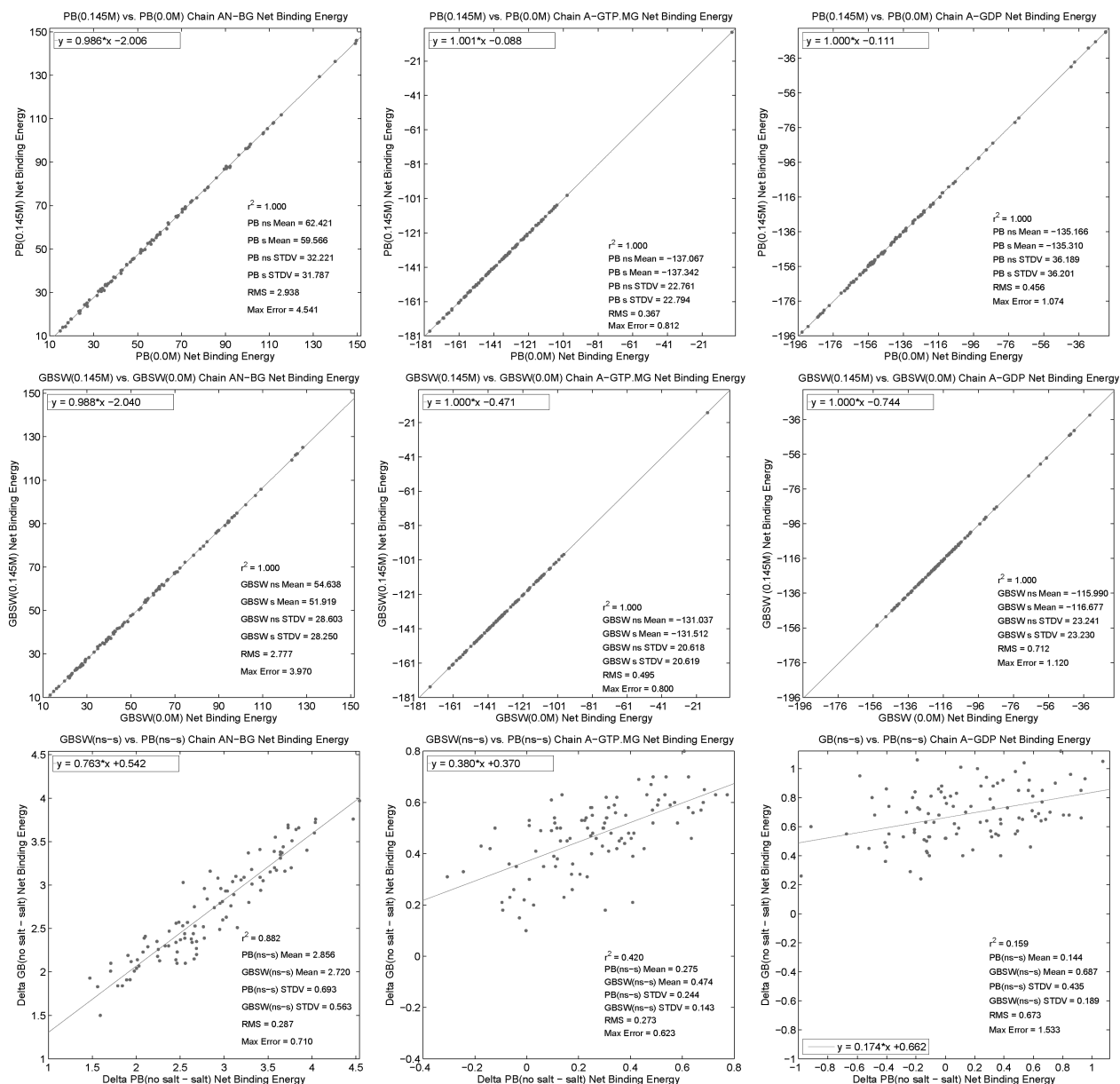


Figure 3.13. Effects of salt on net electrostatic binding free energies. The effects of including a physiological salt concentration of 0.145 M on the net binding free energies computed with both the LPB and GB models are shown for all three systems studied: (left) $G_{\alpha.GDP.G\beta\gamma}$; (center) $G_{\alpha.GTP.Mg}$; (right) $G_{\alpha.GDP}$. Top: The linearized-Poisson-Boltzmann model with 0.145 M NaCl is compared to the Poisson model with 0.0 M ionic strength. Middle: The GB model with salt is compared to the same without. Bottom: The *differences* in net binding energies computed with and without salt are compared between the two models.

3.5.14 Alternative approaches and directions of study

The results presented here were computed with a single implementation of the generalized-Born method (GBSW).⁽¹⁰⁵⁾ In this method, the dielectric boundary between solute

Table 3.3. Salt Effects on Binding Energies.

System	PB ^a		GB ^a		GB vs. PB ^b		
	0.145 vs. 0.0 M		0.145 vs. 0.0 M		0.145 M – 0.0 M		
	rms	max	rms	max	R ²	rms	max
G _{α.GDP} ·G _{βγ}	2.94	4.54	2.78	3.97	0.88	0.29	0.71
G _{α.GTP} ·Mg	0.37	0.81	0.45	0.80	0.42	0.27	0.62
G _{α.GDP}	0.46	1.07	0.71	1.12	0.16	0.67	1.53

^aRoot-mean-square (rms) and maximum absolute difference in net binding energy, in kcal/mol, computed with 0.145 M or with 0.0 M NaCl.
^bCorrelation coefficient, rms, and maximum absolute difference (in kcal/mol) in the effect of salt ($\Delta G_{0.145M} - \Delta G_{0.0M}$) computed with GB or PB.

and solvent is smoothed, while a sharp boundary is used in the Poisson–Boltzmann approach. To evaluate whether this difference contributes significantly to the observed deviations, we repeated all calculations using the GBMV model,⁽⁸³⁾ also within the CHARMM software package. While the results (presented in Appendix A) differ in the details, the deviations seen between GBMV and PB are qualitatively the same as those seen between GBSW and PB (it is worth noting, however, that GBMV gave notably poorly agreement of net binding free energies). GBMV differs from GBSW in a number of ways in addition to the use of an abrupt dielectric boundary, and thus, these results suggest that these deviations are not a peculiarity of one GB implementation.

A number of alternatives are also available,^(80, 104) and these approaches may have somewhat different behaviors. However, the *overall* energetics computed with the two models agree fairly well, and only specific interactions are poorly captured. Thus, it seems unlikely that other approaches would not show similar deviations, if not exactly the same differences.

Deficiencies in GB-based methods have been noted before. For example, Onufriev et. al. have noted the sensitivity of GB-based methods to the method by which effective Born radii are computed, and demonstrated that the use of the so-called “perfect” radii computed from solution of the PB equation results in much better performance.⁽⁸⁸⁾ Unfortunately, the use of these perfect radii involves a solution of the PB equation for each atom in the system; in effect, calculation of the full Φ matrix, which is more than an order of magnitude more costly than performing a residue-by-residue decomposition with PB. While the use of a single reference

structure to compute the effective Born radii may reduce costs, the results presented here clearly demonstrate large variations in group solvation energies with conformation; as these are dominated by individual atomic solvation terms, effective Born radii computed from a single structure would not be able to capture this variation.

Brooks and co-workers have noted that there can be an imbalance between solute-solvent and intersolute interactions with the generalized-Born model,⁽⁸⁷⁾ and as methods are improved to better capture this, it will be worthwhile examining their ability to reproduce individual component energies. Bardhan has also recently described an alternative approach based on a boundary-integral formulation.⁽¹²⁰⁻¹²¹⁾ Interestingly, this method was demonstrated to better reproduce the eigenvectors of the solvation matrix (Φ) than did a generalized-Born approach, and thus, the performance of this method in a system-wide component analysis will also be interesting to evaluate.

3.6 Conclusions

A detailed comparison of the use of both generalized-Born and linearized-Poisson–Boltzmann models in performing a component analysis of the contribution of various groups and energetic terms to binding free energies has been performed. Association of the components of the heterotrimeric G-protein were used as a test case; three binding reactions were considered, including the formation of a protein-protein complex (the G-protein heterotrimer) as well as the binding of guanosine-5'-di- and triphosphate to the monomeric G_α subunit.

The results indicate that in many cases the relative net contributions of individual groups are captured well by the GB model, as measured by the difference in free energy of binding relative to a hydrophobic isosteres what we term mutation free energy. However, this agreement is far from universal, and a significant subset of groups show linear correlations between GB- and PB-computed mutation free energies with slopes much lower than unity. In a few cases, the differences are much more dramatic, with essentially no correlation between the results of the two methods. Individual energetic terms, such as the cost of desolvating a particular group upon binding, tend to be captured much more poorly by the GB-model, as compared to the results from PB. Cancellation of some of this error leads to better results for net contributions, as for total binding free energies. Furthermore, even when relative energetics are captured relatively

well (as indicated by a strong correlation of GB- and PB-computed results), individual values computed with GB often deviate by over 10 kcal/mol.

While universal rules are not clearly present, several observations can be made. First, in general, GB seems to capture PB-based energetics reasonably well for hydrogen-bonded interactions, particularly at buried sites. In contrast, for interactions that act over moderate distances, partially screened by solvent, GB-computed results are often lower in magnitude than those computed by PB; as “action-at-a-distance” interactions of this type have recently been identified as an attractive class of interactions to target in design,⁽¹²²⁻¹²³⁾ the poor performance of GB in this regime should be noted. Finally, GB does not seem to capture the same sensitivity to structural variation of the energetics of buried charged groups as does the PB model. These data clearly indicate a need to take care in interpreting the results of energetic decompositions with a generalized-Born model.

In many cases, however, the primary goal may not be to describe the contribution of every group in a quantitative manner but rather to provide a more qualitative profile of those groups that contribute most to the affinity. In this application, GB is found to be remarkably accurate, and when used to select groups most likely to be significant under a PB-based model, the GB model performs very well. Given the great difference in computational cost of a full system-wide component analysis performed with GB and with PB, the use of GB as an initial screen to identify important components may thus be well merited. This may be followed by Poisson-Boltzmann-based analysis on this focused set of components, providing a reasonable trade-off between accuracy and speed.

Supporting information is available in appendix A.

CHAPTER 4

Structural analysis of the heterotrimeric G-protein in three different states from Long molecular dynamic simulations

4.1 Abstract

Structural differences in the different states of the heterotrimeric G-protein can be expected from the type of nucleotide bound to the α -subunit and the presence—or absence—of the $\beta\gamma$ subunit. These structural differences have been studied with the help of molecular dynamics simulations of the G-protein in its three different states. The resulting trajectories are clustered and compared to quantify structural variations in a systematic way. Of special interest are regions of the α subunit that are known to play an important role in nucleotide exchange and trimer dissociation—two important events in signal transduction through the heterotrimeric G-protein. These important regions correspond to the helical bundle attached to the GTPase domain, the three *switch* loop regions and an additional loop region. A comparative analysis from trajectories gives clues about the function of the G-protein in different states from three different endpoints of its signaling cycle. Differences in structure in these three endpoints can be related to transitions of the protein from one state to another. For example, the helical domain shows a flapping mechanism with a range whose extremes and center are populated differently by the three states of the G-protein. The switch loop regions also behave differently for the different states, with the heterotrimer showing narrower ranges of motions for switch regions one and two, and broader for three and four, compared to the two monomers. These ranges or motions are characterized quantitatively with the help of clusters, and qualitatively by displaying backbone clusters in different colors. This type of clustering analysis is ideal for large protein with multiple subunits that undergo significant changes as in signal transduction. The heterotrimeric G-protein shows clear differences in its trajectories in its three states that indicate that the ranges of motion of these important regions represent a theoretical mechanical model for signal transduction at the molecular level.

4.2 Introduction

Cellular signaling is important for many metabolic processes in multi-cellular organisms such as mammals. The transmission of signals into the cell is modulated by a variety of proteins located inside and near the cell membrane in a concerted mechanism. For example, activation of the trans-membrane G-protein Coupled Receptors (GPCRs) causes structural changes that are passed on to the heterotrimeric G-protein. These structural changes can activate a large number of downstream mechanisms depending on both the type of heterotrimeric G-protein and the tissue or cell in which they occur;(5, 15, 19) these structural changes are partially captured by crystallographic structures of the G-protein in three different states of its signaling cycle, and their study can reveal important structural features with the help of molecular dynamic models and subsequent clustering of their backbones. Results from these theoretical studies reveal important structural features that characterize cellular signaling at a more detailed level of understanding for the heterotrimeric G-protein in particular.

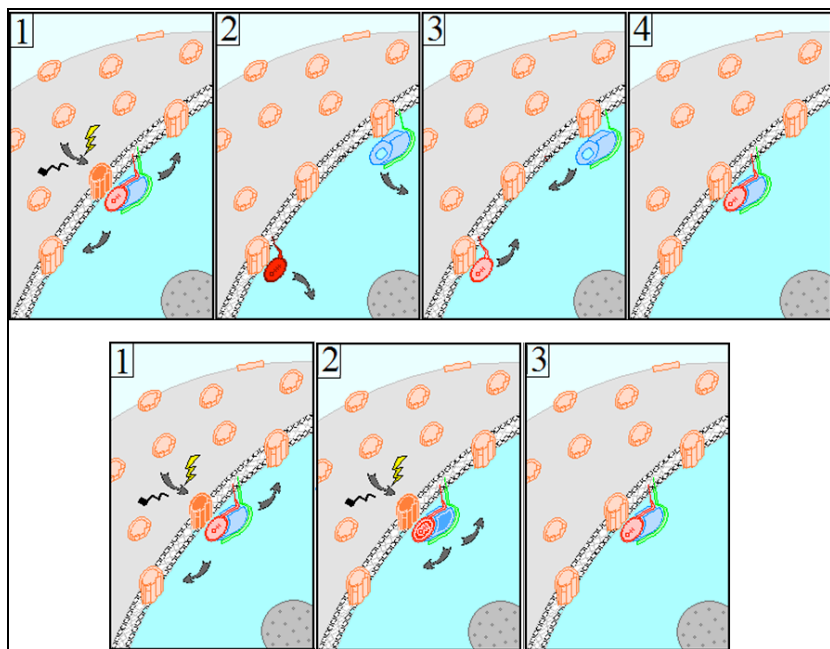


Figure 4.1. Two possible methods of G-protein activation. Top, (1) GPCR activates and signals is passed for nucleotide exchange, (2) the heterotrimer dissociates after nucleotide exchange, (3) the nucleotide is hydrolyzed and the heterotrimer is set for re-association, and (4) the heterotrimer is re-associated and it is set for activation again. Bottom, (1) GPCR activates and signals is passed for nucleotide exchange, (2) activation causes conformational motions that interact with neighboring proteins without dissociation, and (3) when the G-protein goes inactive again, it just remains in its original conformation near the GPCR.

It is important to account for modes of G-protein activation before attempting to relate theoretical models to signaling through heterotrimeric G-protein pathways. The activation of the heterotrimeric G-proteins can lead to two possible ways of signal transduction that are debated in the literature. In one, the heterotrimer completely dissociates and interacts with other downstream effectors. In the second one, the heterotrimer simply undergoes a conformational change between the α and $\beta\gamma$ subunits that activates nearby proteins without fully dissociating.(124) Figure 4.1 shows a scheme of these two possible mechanisms.

Three crystallographic structures that give insights into the structural changes that occur in the heterotrimeric G-protein as it transmits the signal are three $G_{i\alpha 1}$ structures crystallized by Mixon et. al.,(16) Coleman et. al.,(17) and Wall et. al.(18) The first two structures are monomers of $G_{i\alpha 1}$ bound to GDP and GTP, and the last structure includes $G_{\beta 1}$ and $G_{\gamma 2}$ in complex with $G_{i\alpha 1}$ bound to GDP. These three crystal structures correspond to the heterotrimeric G-protein in three different states with identical $G_{i\alpha 1}$ sequence. The crystallographic structures give a lot of insightful structural information, but this information does not capture the full extent of motions within subunits, domains, sub-domains and even side chain conformations. For this reason, computational methods can be used to explore structural differences in the heterotrimeric G-protein in its different states from simulations initiated from crystallographic structures. The simulations provide a time-dependent theoretical model of the motions of the heterotrimeric G-protein in its three different states solvated in a box of waters with Na^+ and Cl^- ions at a concentration equal to human physiological conditions.

A comparison between motions in simulation to experimental observations can be done by plotting temperature B-factors. These factors give a measure of mobility of each atom. B-factors are obtained from x-ray diffraction and are also calculated from Root Mean Square Fluctuation (RMSF) calculations from the simulations. Equation 4.1 is used to convert RMSFs of atom i to its corresponding B-factor value.(125)

$$B_i = \frac{8\pi^2}{3} (RMSF_i)^2 \quad (4.1)$$

Making inferences that can be related to function from molecular dynamic simulations can be done with the help of clustering algorithms. There are many clustering algorithms that can

be used in a variety of scenarios depending on the characteristics of the problem. There is no one clustering algorithm that work across different data mining problems. Even when the data that is available for searching corresponds to structural information from molecular dynamics simulations, different cluster algorithms for the same data set could give different results as explored by Shao et. al.(126) For this reason, a clustering algorithm for the problem in hand needs to be selected to facilitate interpretation of results.

Many clustering algorithms have been inspired by the need of organizing structures generating by computational folding studies of small peptides. Kerpen et. al. applied a nonhierarchical clustering algorithm based on a self organizing neural network on a pentapeptide.(127) Satoh et. al. (128) use NMR structures from a 10 residue peptide to cluster conformations from molecular dynamics within a cutoff of 1 Å; in a second round of clustering pair-wise C α RMSD were “superclustered” with a single linkage algorithm and a 0.4 Å cutoff. These clustering algorithms do not specify a specific alignment reference, since in folding of small peptides the structures change more and lack a well defined core structure as in the heterotrimeric G-protein.

In this work, rigid sections of the heterotrimeric G-protein are used for alignment before clustering. The regions used for clustering are the size of small peptides but are constrained in both ends by secondary structure regions of the G-protein that are more rigid. Clustering is then reduced to grouping structures that pivot around a reference frame. Meaningful interpretation of this type of clustering is possible as long as the reference frame is the same in the three systems; that requires alignment of the three trajectories to the same—and only one—structure.

The structural differences between the three systems are due to the presence of GDP, GTP or $\beta\gamma$ in the inactive monomer, active monomer and the inactive heterotrimer respectively. Regions of the heterotrimeric G-protein that are the most different in its three different states will be analyzed, grouped and displayed by clusters obtained from 2D-RMSD differences between frames from simulations. This analysis of the heterotrimeric G-protein will identify structural determinants of function from theoretical studies that could be important for signal transduction in living organisms; thus, expanding on structural information that can be derived from crystal structures. The following methodology for clustering is suitable for larger molecules with well define core structures and multiple loop regions that are attached to sequence-separated regions

of the protein. This chapter organizes the structures in simple and intuitive way that can be used as a guide for analyzing side chain conformations in the next chapter.

4.3 Methods

4.3.1 Structure preparation

Initial heavy-atom coordinates were obtained from the Protein Data Bank (1gia, 1gdd, and 1gp2).^(107, 129) While $G_{\alpha 1}$ has the exact same primary structure in the three systems, not all of the residues in the sequence diffracted the X-rays to reveal the positions of the atoms. Different sections of $G_{\alpha 1}$ were missing in each one of the three crystal structures. These sections were completed by homology modeling after the available common structural information was aligned with the program PROFIT.⁽¹³⁰⁾ Missing sections from the alignment were concatenated to the PDB files in the three systems; thus, making $G_{\alpha 1}$ sequence identical in the three systems. Protonation states were chosen with the REDUCE program, as were the orientations of amides and imidazole rings;⁽¹⁰⁸⁾ this resulted in a single protonated histidine in the subunit of the trimer (1gp2), with all other histidines in a neutral state, all aspartates and glutamates negatively charged, and all lysines and arginines positive. Hydrogen atom positions were determined with the HBUILD module ⁽¹⁰⁹⁾ of the CHARMM ^(110, 131) computer program.

4.3.2 Molecular dynamics

All-atom, explicit solvent molecular dynamics simulations were performed using the CHARMM ^(110, 131) and NAMD ⁽¹¹¹⁾ programs with PARAM22/27 parameters;⁽¹¹²⁻¹¹³⁾ system setup and post processing were done with CHARMM, while production dynamics simulations were run using NAMD. Each complex was placed in a pre-equilibrated box of TIP3P waters,⁽⁴⁵⁾ with a minimum of 10 Å between any solute atom and the box edge in all directions. Random water positions were replaced with enough sodium and chloride ions to yield physiological (145 mM) ionic strength; the sodium to chloride ratio was adjusted from unity to provide a system with zero net charge. Periodic boundary conditions were applied to minimize edge artifacts, and particle-mesh Ewald (PME) summation was used for long-range electrostatic interactions; short-range interactions were cut off at 12 Å. With the previous initial starting conditions, a time step of 2 femtoseconds was used to run the simulations for 608 ns of simulation time, after 240 steps of minimization and 200 ps of equilibration using Langevin dynamics implemented in NAMD.

4.3.3 Preparation and alignment of trajectories for comparison

The resulting trajectories were stripped the waters and ions for structural analysis. G_{α} was selected for alignment because this is the only monomer common to the three structures. G_{α} was divided into regions that were selected to separate parts of the structures with low and high temperature factors—or B-factors—from the crystallographic structures. Figure 4.2 shows a secondary and quaternary structure division of the regions used for fitting all the frames in all the trajectories to a reference frame. Table 4.1 defines the regions of the sequence used for dividing the α subunit.

Table 4.1. Definition of regions used for trajectory fitting and structural analysis of G_{α} .^a

Region	Sequence	Color
N-terminus	6 - 34	Purple
GTPase	35 – 40, 50 – 59, 187 – 201, 221 – 231, 244 – 281, 296 - 326	Green
P Loop	41 - 49	Blue
Helical	60 - 178	Red
Switch 1	179 - 186	Cyan
Switch 2	202 - 220	Cyan
Switch 3	232 - 243	Cyan
Switch 4	282 – 295	Cyan
C-terminus	327 - 354	Blue
Nucleotide	355	Orange

^aThe sequence regions can be seen in Figure 4.3.

4.3.4 Two dimensional root mean square deviations (2D-RMSD)

The root-mean-square deviation (RMSD) of each molecular dynamics frame in the simulation to every other frame was organized in a two dimensional matrix. These matrices of RMSDs were calculated from trajectories that had been fitted using the more rigid sections of the GTPase domain described in the previous method section. The 2D-RMSD is usually displayed as a square matrix whose numerical values are color coded to ease the visual identification of conformational transition for the section of the protein analyzed. This information is useful to assess the relative change of each structural frame to every other frame in the simulation; thus, showing an all-to-all frame comparison of the systems. This type of analysis is sensitive to the

regions used for alignment, it is important that this region is consistent for the three systems in order to make the analysis comparable.

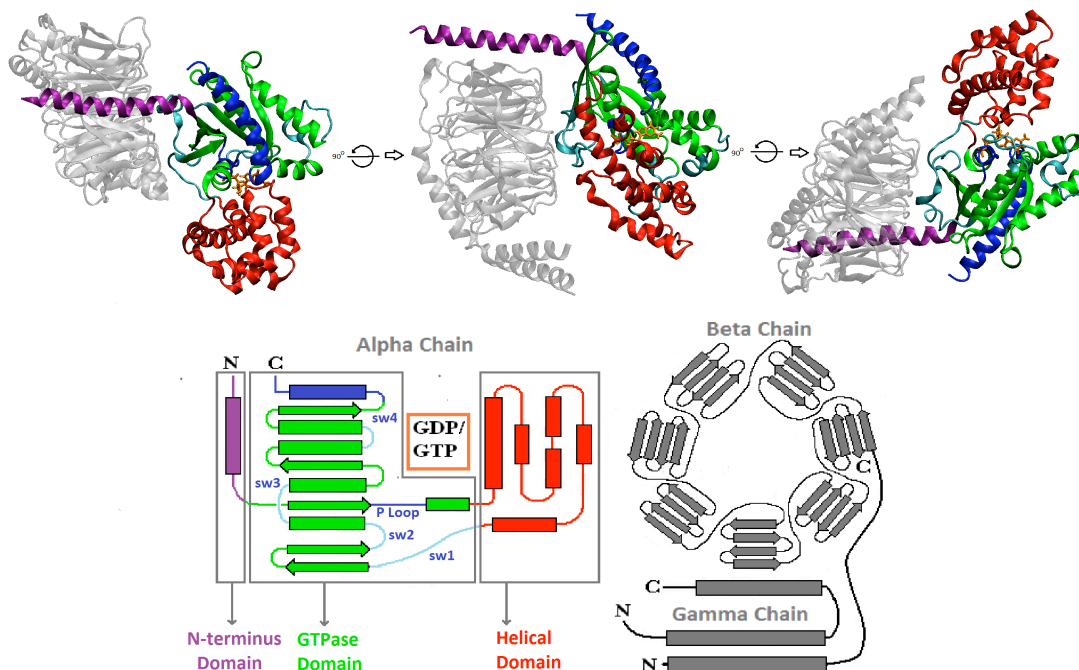


Figure 4.2. Subdivision of G_α for trajectory fitting and structural analysis. Top, quaternary structure of the heterotrimer. Bottom, secondary structure representation of the heterotrimer with helical regions represented by rectangles, loop regions by lines, and beta sheet regions by arrows. Both figures are colored identically to show where regions are located in the protein sequence and in the structure. Regions used for fitting all frames in the trajectories are colored green, and the other colors are used to identify regions of interest: the N-terminus domain in purple; the helical domain in red; the nucleotide in orange; the C-terminus domain in blue; and loop regions in cyan and blue. Only loop regions in cyan are analyzed in this chapter. The $\beta\gamma$ subunits are colored in grey.

4.3.5 Dendrograms from two dimensional RMSDs

It is possible to measure an Euclidean distance from the RMSD of each frame in the simulation to every other frame. Each row or column in the 2D-RMSD can be considered a 608 dimensional vector, and the distance between each vector can be measured by Equation 4.2, where j and k correspond to vectors formed by two different frames, identify as f , and n is the number of frames, or dimensions.

$$D_{j,k}^n = \sqrt{(f_j^1 - f_k^1)^2 + (f_j^2 - f_k^2)^2 + \dots + (f_j^{608} - f_k^{608})^2} \quad (4.2)$$

The distance information between all frames can be organized in a dendrogram that shows the grouping of frames, and groups of groups of frames, closest to each other. In order to get the dendrogram, the 2D-RMSDs are input into the MATLAB (132) function *pdist* with the Euclidian parameter. This function outputs a matrix with the distance between all frames that is then input into the MATLAB *linkage* function with the second parameter set to average. The resulting N-1 by 3 matrix is input into the *dendrogram* function for display; N represents the dimension of the 2D-RMSD matrix. The dendrograms show the size of each of these clusters displayed in different colors, and the envelope plots next give a reference of how these clusters' distances relate to Angstroms. The clusters do not give information about the order in which the clusters alternate as the simulation goes on; that information is organized right below it by showing the clusters that are populated throughout the simulation, followed by a bundle of all backbone structures in the simulation colored according to the cluster they belong.

4.3.6 Average and maximum cluster RMSDs

In the dendrogram from a simulation trajectory with N snapshots there are N-1 possible ways to assign clusters based on a distance cutoff. For clusters formed by more than one snapshot, their average RMSD was calculated, and then cluster averages are calculated and joined with a node to the closest cluster until all possible clusters are linked. The corresponding threshold for each cluster is plotted as a function of distance in angstroms; this threshold is defined as the smallest one that gives the given number of clusters from 1 to N-1. Some thresholds of special interest are those corresponding to a minimum average separation of 3, 2.5 and 2 Å between them. In other words, thresholds for which the separation of the clusters it defines is less at least 3, 2.5 and 2 Å. The results are then used as a guide to select a threshold that would give a reasonable number of clusters for each section of the protein; this allows a detailed description of differences in backbone structural snapshot bundles among the three systems.

4.3.7 Clustering of regions from separate simulations

It is possible to cluster the frames from the three simulations by doing the exact same procedure as for a single simulation. For that purpose, the matrices for each one of the regions in the previous section are used, in addition to matrices that measure 2D-RMSDs from frames in separate trajectories. Distance matrices between frames in the same trajectory were placed in the diagonal and off the diagonal distance matrices across different systems. The three systems gave

a 3 by 3 block matrix with the upper diagonal transposed for symmetry purposes. Clustering analysis was done on this 1824 by 1824 symmetric matrix. The matrix was then rearranged to coincide with the ordering of the frames from the three simulations as they were grouped in the dendrograms. The system to which each frame belonged in the rearranged matrix was then used to plot how frames from different simulations were group by proximity. This is presented in the middle of Figures 4.6, 4.8, 4.10, 4.12 and 4.14 with three values on the y-axis that corresponds to each of the systems. Finally, bundled backbone representations from the trajectories are shown next to the clustering, showing proximity patterns between conformational space sampled by the three systems and the clustering from 2D-RMSDs.

4.4 Results and discussion

Structural analysis was first done on the large domain regions of the α subunit, followed by important loop regions considered important in structural transitions in G-protein signaling. The N-terminus was not given a full analysis because results can only be interpreted if the cell membrane is present. The GTPase domain is presented only with 2D-RMSD because of its lack of mobility and its use for fitting. The switch loop regions and the helical region are given a full clustering analysis.

4.4.1. Motions between sub-domains

4.4.1.1 N-terminus domain

It is difficult to compare 2D-RMSDs, dendrograms and clustering for the N-terminus domain with the GTPase domain used for alignment. This region has modifications that penetrate the cell membrane, and it is missing from the crystallographic structure of $G_{\alpha,GTP,MG}$. Missing structural information was added from available structural information from $G_{\alpha,GDP}$, making the N-terminus in the monomers relatively similar in the simulations; however, in the heterotrimer, the N-terminus is stretched along the $\beta\gamma$ subunit. For these reasons it is difficult to make any possible interpretations about the role of this subdomain in signal transduction from molecular dynamics simulations. Figure 4.3 shows how this region varies for the three systems. It is interesting to see clear differences between the active and inactive monomer, blue and yellow in Figure 4.3, despite having identical starting structural information. The differences are due to initial crystallographic differences in the GTPase domain and the presence of a different

nucleotide in the monomers. It is interesting to note that in this long simulation the same region explores evidently different conformations. It is not clear that this is due to lack of convergence or that these regions explore different conformational *wells*.

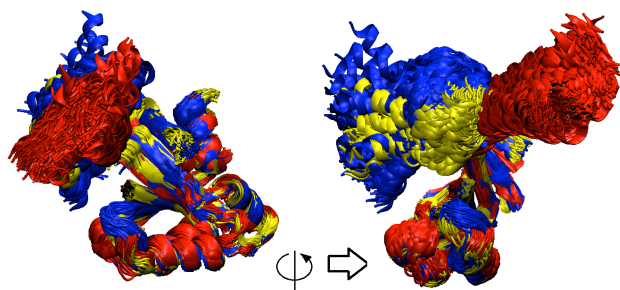


Figure 4.3 Variation in N-terminus structure. Difference in N-terminus domain orientation relative to the GTPase fitting domain for the heterotrimeric G-protein in its three different states: red for $G_{\alpha.GDP.G\beta\gamma}$; yellow for $G_{\alpha.GTP.MG}$; and blue for $G_{\alpha.GDP}$.

4.4.1.2 GTPase domain

This domain ‘harbors’ the nucleotide. Important differences between the three systems are due to the loop regions in this domain and are excluded from the alignment. Beta-sheets and helical sections are used for alignment so that other regions of the protein can be compared between the three systems. Figure 4.4 shows this comparison between systems in the form of 2D-RMSD matrices and their corresponding ranges of mobility in angstroms next to them, with the superposition of backbone structures to the right; $G_{\alpha.GDP}$ has a maximum RMSD between frames of 1.97 Å, $G_{\alpha.GTP.MG}$ of 1.84 Å and $G_{\alpha.GDP.G\beta\gamma}$ of 2.32 Å. The nucleotide is not exclusively surrounded by this region of the protein, but it has residues that allow it to interact favorably with the phosphate groups primarily. The nucleotide is also in contact with switch region one and the α -helical domain. Clustering did not evidenced differences that could be structurally differentiated and is not presented for this region of the protein.

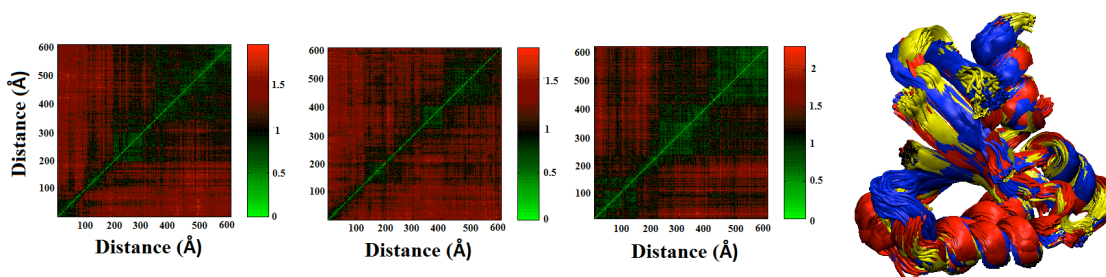


Figure 4.4. Variation in GTPase domain structure. From left to right, 2D-RMSDs for $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$, and alignment of the GTPase domain used for fitting the trajectories for the heterotrimeric G-protein in its three different states: red for $G_{\alpha.GDP.G\beta\gamma}$; yellow for $G_{\alpha.GTP.MG}$; and blue for $G_{\alpha.GDP}$.

4.4.1.3. The α -helical domain

The nucleotide is partially surrounded by residues in the α -helical domain. It has been suggested that nucleotide exchange occurs when the GTPase domain and the α -helical domain separate following GPCR activation.^(21, 133) Figure 4.5 shows a systematic analysis of the α -helical domain's range of motions in the three systems. As a first step, the magnitude range of these motions can be seen in the scales to the right of the 2D-RMSDs; they show that $G_{\alpha.GTP.MG}$ has a maximum RMSD of 10.18 Å, while $G_{\alpha.GDP.G\beta\gamma}$ goes up to 12.98 Å and $G_{\alpha.GDP}$ to 12.54 Å. This suggests that the α -helical region of $G_{\alpha.GTP.MG}$ is generally more rigid in terms of mobility relative to the GTPase domain by ~ 2 Å.

After clustering the 2D-RMSDs, a minimum threshold distance between frames was selected to give a reasonable number of clusters in the three systems. This selection was guided by a relationship between distance and angstroms in the plots next to the dendrograms in Figure 4.5. The selection of this threshold for clustering is also explained in Section 4.3.6, and was chosen to be 38 in order to group the blocks in the 2D-RMSDs; this threshold separates the trajectories of $G_{\alpha.GDP}$ into 6 clusters, $G_{\alpha.GTP.MG}$ into 5, and $G_{\alpha.GDP.G\beta\gamma}$ into 6. One fewer clusters suggest that $G_{\alpha.GTP.MG}$ is more rigid. This is supported by having a predominant cluster in yellow for 389 snapshots, followed by a smaller cluster (green) with 88 frames and the remaining 23 frames distributed among the remaining three clusters. $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$ do not show such concentration of frames in one single cluster with 257 and 267 frames for the bigger clusters respectively.

A top view of the clusters at the bottom of Figure 4.5 shows that the motions are generally sideways as if the domain flaps relative to the GTPase domain, with clusters clearly separated from one extreme of these motions to the other. The clusters that are colored magenta are the bigger outliers in the three systems, and show a larger separation in the dendrograms of Figure 4.5. For $G_{\alpha.GDP}$, this cluster is found in the middle of the simulation—implying a rare sampling of a conformation. For $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$ this cluster is found at the beginning of the simulations—implying a strained conformation in the crystal structure. In the three systems, the magenta cluster shows some regions from which it protrudes from the bundle of clusters. While this protuberance is not very noticeable, it is relevant considering it corresponds to a few snapshots of the total 608; however, it was difficult to pinpoint to a single loop or helical region

as the cause for the large distance of this cluster to the remaining clusters. The bottom row of Figure 4.5 shows that the differences in clusters across systems are due to a large motion of the entire domain that is similar to a flapping mechanism relative to the GTPase domain. The extreme ends of this flapping mechanism correspond to $G_{\alpha.GDP}.G_{\beta\gamma}$ and $G_{\alpha.GDP}$, with $G_{\alpha.GTP.MG}$ sampling an intermediate space in between.

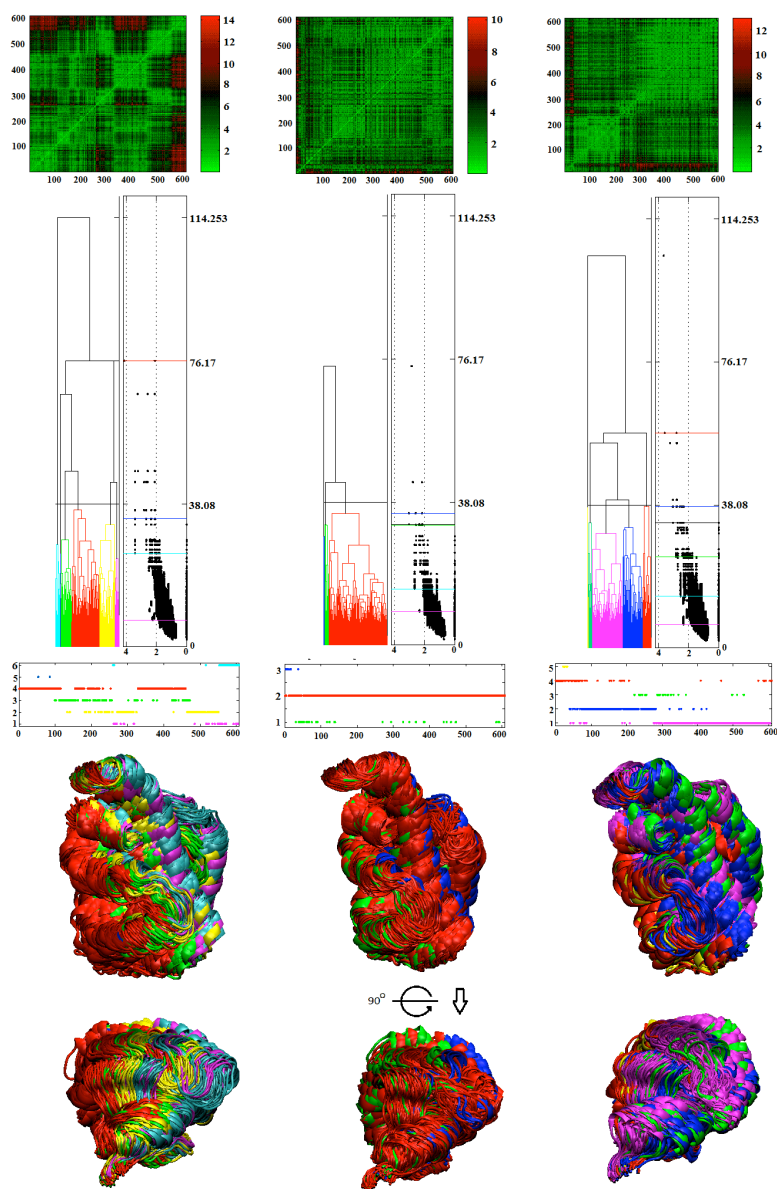


Figure 4.5. Structural variation in the α -helical domain. The columns correspond to α -helical domain ordered from left to right: $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP}.G_{\beta\gamma}$. The top row corresponds to 2D-RMSDs deviations between frames in each trajectory. The second row shows the dendrograms that group distances between groups of frames. Clusters are defined and colored with a threshold distance of 38. Average distances between frames are plotted against the corresponding average distances in angstroms in the envelope plots next to the dendrograms. As a guide to relate thresholds to angstroms, color coded thresholds are shown for the mean average of frame averages below 3 (red), 2.5 (blue) and 2 (green) Å. Also, for maximum average of frame averages below 3 (black), 2.5 (cyan) and 2 (magenta) Å. The following row shows how the clusters from the dendrograms alternate through the simulation. The structures in the row below show how the clusters separate structures in the simulation from two different points of view.

It is possible to cluster all frames corresponding to the molecular dynamics trajectories of the three systems. This is done by aligning the three 2D-RMSDs from Figure 4.5, and calculating 2D-RMSDs between different trajectories of the three systems off the diagonal. The resulting

3 x 3 block matrix can be clustered in a similar way as it is done for individual systems. This is done in Figure 4.6 where the three systems are clustered. Three clusters emerge from the dendrogram if a threshold between 113.34 and 152.27 is selected. Under this threshold, all of the systems overlap with each other; $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$ overlap sporadically, and $G_{\alpha.GTP.MG}$ overlaps more often with the other two. The range of this motion is 18 Å with a maximum distance between clusters of 191.59.

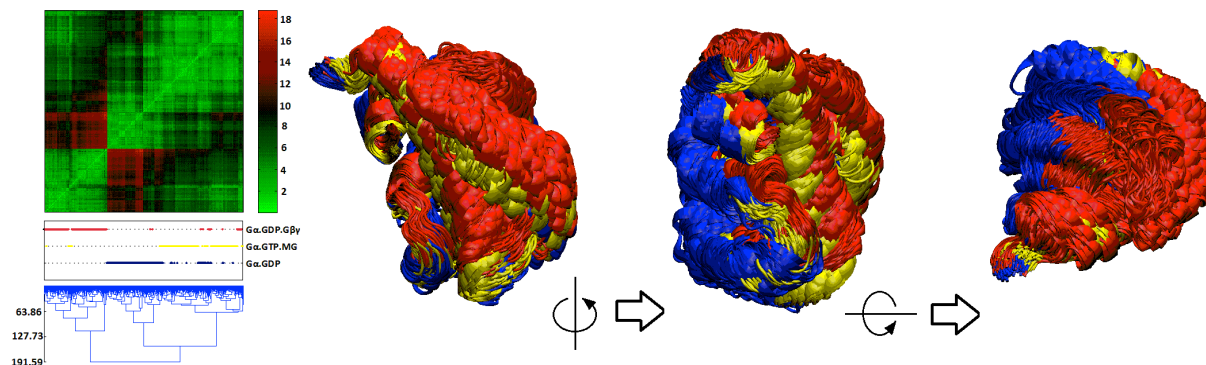


Figure 4.6 Variations in α -helical structure are common in all simulations. α -helical domains clustering of combined 2D-RMSDs from different trajectories. Left, 2D-RMSDs from Figure 4.5 are used to form a 3 by 3 matrix for clustering. Clustering shows that $G_{\alpha.GTP.MG}$ samples conformations closer to the other two systems. $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$ seem to sample distant conformations that do not overlap, as seen in the dendrogram and the coloring of the clusters according to system. Right, in the same coloring scheme, three angles of the α -helical region show that indeed $G_{\alpha.GTP.MG}$ seems to be in the middle, yellow, of the other two systems, $G_{\alpha.GDP.G\beta\gamma}$ in red and $G_{\alpha.GDP}$ in blue. These motions relative to the GTPase domain act like a flapping mechanism suggesting different degrees of mobility when different nucleotides are present and when the $\beta\gamma$ subunit is present. This suggests a large structural bias between subdomains in the release, exchange and hydrolyzation of the heterotrimeric G-protein in its different states.

4.4.2. Motions in switch regions.

Four loop regions will be considered for structural analysis in this paper. Three of those are the switch regions referred by crystallographic papers to be important in heterotrimeric G-protein. A fourth region is a loop region located far from the $\beta\gamma$ subunit, the nucleotide and the α -helical region. This region is not considered to be very important for the heterotrimeric G-protein to function, yet it takes different conformations in the three different states that suggest some allosteric reaction to the presence of a different nucleotide or $\beta\gamma$.

4.4.2.1. Switch region one

This switch region is located between the GTPase and the α -helical sub-domains, and it connects the two sub-domains through a region that is closer to the nucleotide. The other section of the protein that connects the two main sub-domains is not close to the nucleotide and connects two long helical regions in both domains. This region varies largely among the three systems. The ranges of motions are initially described from 2D-RMSD plots for the three systems. Figure 4.7 shows that $G_{\alpha.GDP}$ has three main blocks clearly defined; a small one at the beginning, followed by a large block throughout most of the simulation, and smaller one towards the end. $G_{\alpha.GTP.MG}$ has two predominant blocks, though the second block, towards the end, may be formed by three, or more, smaller clusters closer to each other than to the first one. Finally, $G_{\alpha.GDP.G\beta\gamma}$ has many blocks indicating multiple clusters close to each other in contrast to the large block transitions in the two monomers. The RMSD ranges for the two monomers are 7.81 Å for $G_{\alpha.GDP}$ and 8.84 Å for $G_{\alpha.GTP.MG}$, for the heterotrimer the maximum RMSD is 4.40 Å.

The average distances between frames in the clusters can be observed in the dendrograms below. A threshold of 28 was selected because it gives a reasonable number of clusters for the three systems. This threshold divided this region in 5 clusters for $G_{\alpha.GDP}$ and $G_{\alpha.GTP.MG}$. As for $G_{\alpha.GDP.G\beta\gamma}$, it gives only one cluster under this threshold, indicating its lack of motion compared to the monomers. While the two monomers have the same number of clusters, the maximum distance between their clusters is different. $G_{\alpha.GDP}$ has all of its clusters closer together and separated by a maximum vector distance of 51.23; $G_{\alpha.GTP.MG}$ has a larger maximum separation of 99.40 between clusters. The corresponding distance in angstroms for these two clusters separations is very close considering that the average frame vector distance is almost twofold. For $G_{\alpha.GDP}$ the largest separation between clusters is equivalent to 2.90 Å, and for $G_{\alpha.GTP.MG}$ the largest separation is 3.23 Å. This discrepancy in angstroms between cluster distances from vectors and angstroms occurs because they are simply averages calculated from frames that could be spread differently: the standard deviation for $G_{\alpha.GDP}$ is 1.18 Å and for $G_{\alpha.GTP.MG}$ is 2.02 Å, a larger spread between clusters in the active monomer compared to the inactive monomer. The nucleotide interacts strongly with some of the residues in this region and it is a factor in the difference in motion.

In the active monomer, the nucleotide coordinates a magnesium ion with the help of its additional phosphate group. Additionally, Thr 181 interacts with the magnesium ion from the beginning of the simulation. Dissociation of this residue from the magnesium ion is an event that coincides with a larger range of conformations that in turn generates clusters of backbone structures that are farther from the original one.

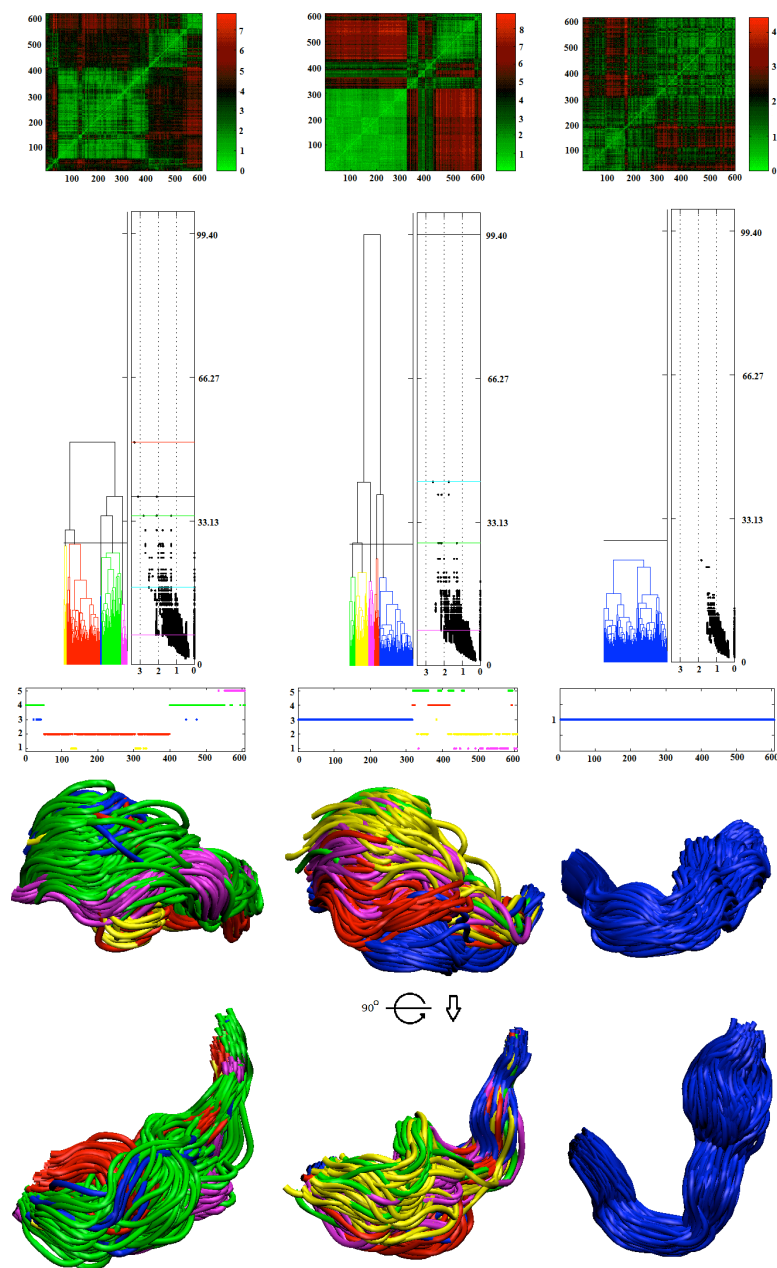


Figure 4.7. Structural variation in switch region one.

The columns correspond to α -helical domain ordered from left to right: $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G_{\beta\gamma}}$. The top row corresponds to 2D-RMSDs deviations between frames in each trajectory. The second row shows the dendrograms that group distances between groups of frames. Clusters are defined and colored with a threshold distance of 28. Average distances between frames are plotted against the corresponding average distances in the envelope plots next to the dendrograms. As a guide to relate thresholds to angstroms, color coded thresholds are shown for the mean average of frame averages below 3 (red), 2.5 (blue) and 2 (green) Å. Also, for maximum average of frame averages below 3 (black), 2.5 (cyan) and 2 (magenta) Å. The following row shows how the clusters from the dendrograms alternate through the simulation. The structures in the row below show how the clusters separate structures in the simulation from two different points of view.

The difference in mobility between the two monomers for this switch region can be better described from how different clusters are sampled through the simulation. $G_{\alpha.GDP}$ appears to start at a strained conformation from the crystallographic origins as suggested by cluster 3 and 4 in green and blue, before transitioning to cluster 2 in red. $G_{\alpha.GTP.MG}$ starts in a more stable conformation corresponding to cluster 5 in green. The active monomer, $G_{\alpha.GTP.MG}$, remains in this state until after 316 nanoseconds into the simulations when that cluster of conformations is never sampled again. $G_{\alpha.GDP.G\beta\gamma}$ is very rigid judging by the smaller range of RMSD that groups all frames in one cluster below the threshold of 28.

Figure 4.8 groups all frames in the three systems clustered by proximity with corresponding backbone bundling colored for each state of the heterotrimeric G-protein. Clusters red and blue—corresponding to $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$ respectively—are separated in two extreme conformations with the yellow cluster— $G_{\alpha.GTP.MG}$ —spanning conformations between the other two clusters.

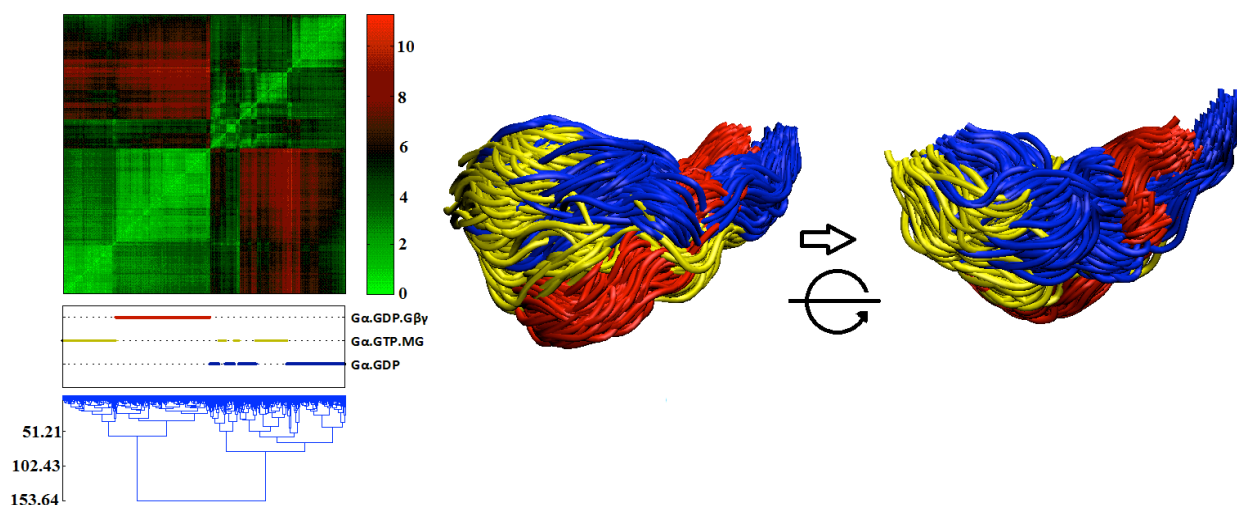


Figure 4.8 Structural variations in switch region one are similar for the monomers. Switch region one clustering of combined 2D-RMSDs from different trajectories. Left, 2D-RMSDs from Figure 4.7 are used to form a 3 by 3 block matrix for clustering. Clustering shows that $G_{\alpha.GTP.MG}$ samples conformations closer to the other two systems. $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$ seem to sample distant conformations that do not overlap. Right, in the same coloring scheme, two angles of switch region one show that indeed $G_{\alpha.GTP.MG}$ seems to be in the middle, yellow, of the other two systems, $G_{\alpha.GDP.G\beta\gamma}$ in red and $G_{\alpha.GDP}$ in blue.

4.4.2.2 Switch region two

This region is not in direct contact with the nucleotide in any of the structures. Some of its residues do interact with other residues that in turn interact with the nucleotide. In the

heterotrimer, this region is in contact with the $\beta\gamma$ subunit. Backbone structural differences for the three systems show that this region is between the nucleotide and the protein-protein binding interface formed by α and β . Figure 4.10 shows the 2D-RMSD plots for this region. This region shows more mobility in the inactive monomer, $G_{\alpha.GDP}$, with a maximum distance between frames of about 5.08 Å. The active monomer, $G_{\alpha.GTP.MG}$, has a range of RMSDs that goes up to 3.41 Å, and the heterotrimer $G_{\alpha.GDP.G\beta\gamma}$ has a range of 3.94 Å. This region did not show any electron diffraction in the crystals structures of the inactive monomer, and was completed with sections of the active monomer. It is important to note that molecular dynamic simulations show an increased mobility for the inactive monomer that the crystal structure suggests from the lack of refraction.

The 2D-RMSDs were clustered with a threshold of 15 to give a reasonable number of clusters in the three systems, as seen in Figure 4.9, and $G_{\alpha.GDP}$ gives seven clusters, $G_{\alpha.GTP.MG}$ gives only two, and $G_{\alpha.GDP.G\beta\gamma}$ gives five clusters. A slightly lower threshold for $G_{\alpha.GDP}$ would divide the red cluster in two, for a total of eight clusters, and no additional clusters for $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$. The reason for not using a threshold lower than 15 is that $G_{\alpha.GDP}$ already has many clusters, and an additional one would make it harder to interpret the differences in structures.

The structures grouped by clustering for $G_{\alpha.GDP}$ are not very easy to distinguish from the color coded bundles of snapshots in Figure 4.9. For this same system, it can be seen from the distributions of clusters in time in the figures above, that four of the clusters, black, green magenta and blue, are sampled only for a few snapshots in the simulations. Only the yellow red and cyan clusters are sampled for significant periods of time, but their corresponding backbone structures don't merit any further subdivisions of these clusters by a lower threshold. The two clusters in $G_{\alpha.GTP.MG}$ also have evident differences, with the red cluster deviating sampling towards the end of the simulation. Finally, $G_{\alpha.GDP.G\beta\gamma}$ shows that the alpha helix structure of this region is interrupted in the middle; also, the region in this system appears more compact or rigid; this is supported by more evenly distributed clusters in the dendrograms shown in Figure 4.9. In summary, $G_{\alpha.GDP}$ is less rigid than $G_{\alpha.GTP.MG}$, and this one in turn, is more compact than $G_{\alpha.GDP.G\beta\gamma}$; thus, differences in the nucleotide influence this region even though there is no direct

interaction with them, $\beta\gamma$ also must play a role in the interruption of the alpha helix structure of this region in the heterotrimer.

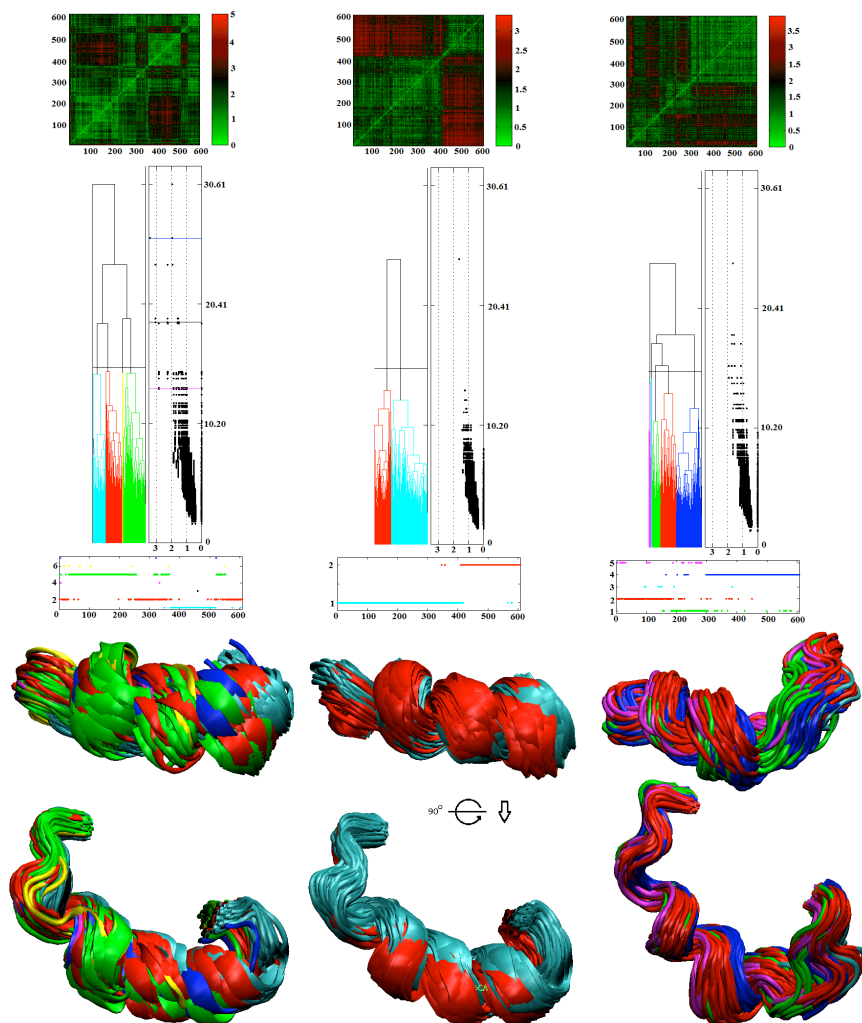


Figure 4.9. Structural variation in switch region two. The columns correspond to α -helical domain ordered from left to right: $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$. The top row corresponds to 2D-RMSDs deviations between frames in each trajectory. The second row shows the dendrograms that group distances between groups of frames. Clusters are defined and colored with a threshold distance of 15. Average distances between frames are plotted against the corresponding average distances in angstroms in the envelope plots next to the dendrograms. As a guide to relate thresholds to angstroms, color coded thresholds are shown for the mean average of frame averages below 3 (red), 2.5 (blue) and 2 (green) \AA . Also, for maximum average of frame averages below 3 (black), 2.5 (cyan) and 2 (magenta) \AA . The following row shows how the clusters from the dendrograms alternate through the simulation. The structures in the row below show how the clusters separate structures in the simulation from two different points of view.

Figure 4.10 shows all frames in the three systems clustered by proximity with corresponding backbone bundling colored for each state of the heterotrimeric G-protein. Clusters

yellow and blue—corresponding to $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP}$ respectively—maintain an uninterrupted alpha helical secondary structure better than the cluster corresponding to $G_{\alpha.GDP.G\beta\gamma}$, in red. The range of RMSD for this region is 7.74 Å.

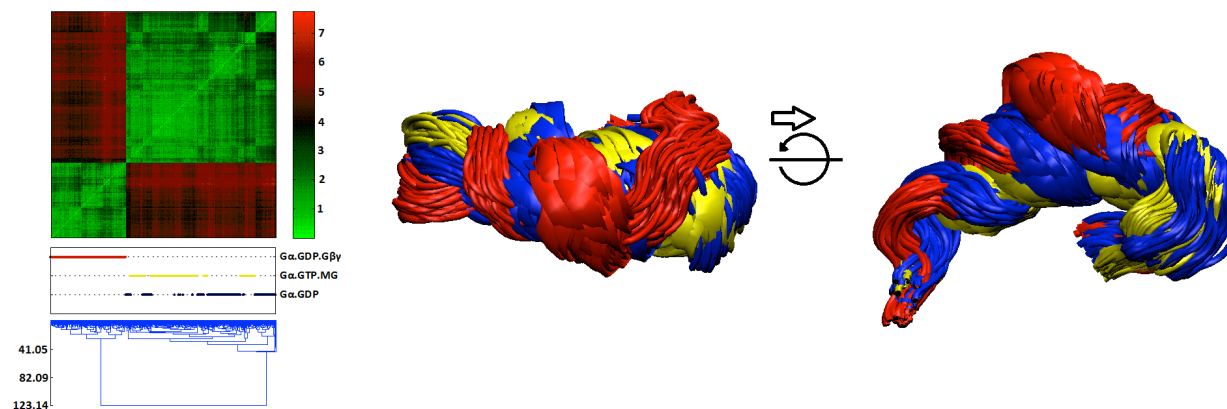


Figure 4.10. Switch region two adopts unique structure in the trimer. Switch region two clustering of combined 2D-RMSDs from different trajectories. Left, 2D-RMSDs from Figure 4.9 are used to form a 3 by 3 block matrix for clustering. $G_{\alpha.GDP.G\beta\gamma}$ sample distant conformations that do not overlap with the two monomers, most likely due to the presence of $\beta\gamma$. Right, in the same coloring scheme, two angles of switch region one show that $G_{\alpha.GDP.G\beta\gamma}$ has a bundle of backbone structures that clearly separate from the two monomers, $G_{\alpha.GDP}$ and $G_{\alpha.GTP.MG}$, in blue and yellow respectively.

4.4.2.3. Switch region three

This region is located below the nucleotide, opposite to the N-terminus C-terminus of the α -subunit relative to the cell membrane. This loop region is in an intermediate position within reach of interaction with the other two loop regions, the α -helical domain and the p-loop region that *harbors* the phosphate groups of the nucleotide. Because of its locations, its importance appears to be to mediate with other important regions of the heterotrimeric G-protein.

A particular feature of this region is the presence of three glutamate residues that can reach and interact with Arg 205 in switch region 2, Gln 147 in the α -helical domain and Lys 180 in switch region 1. Figure 4.11 shows that this switch region has a large range of motions in the inactive monomer, followed by the active monomer and then by the heterotrimer. The heterotrimer seems to start in a strained position and then it relaxes into two large conformations for the remaining of the simulations.

Clustering of the 2D-RMSD was done from a judicious selection of a threshold of 25 for the three systems as seen in Figure 4.11. This threshold forms four clusters in $G_{\alpha_{\text{GDP}}}\cdot G_{\beta\gamma}$, of which three are small. This suggests that the main blocks seen in the 2D-RMSD correspond to similar structures and not transitions to new conformations. The smaller clusters are brief departures from the main conformation. $G_{\alpha_{\text{GTP.MG}}}$ has three clusters that include a larger cluster, a medium second cluster, and only one small transitional cluster. $G_{\alpha_{\text{GDP}}}\cdot G_{\beta\gamma}$ has in turn five clusters with two large ones and three small ones.

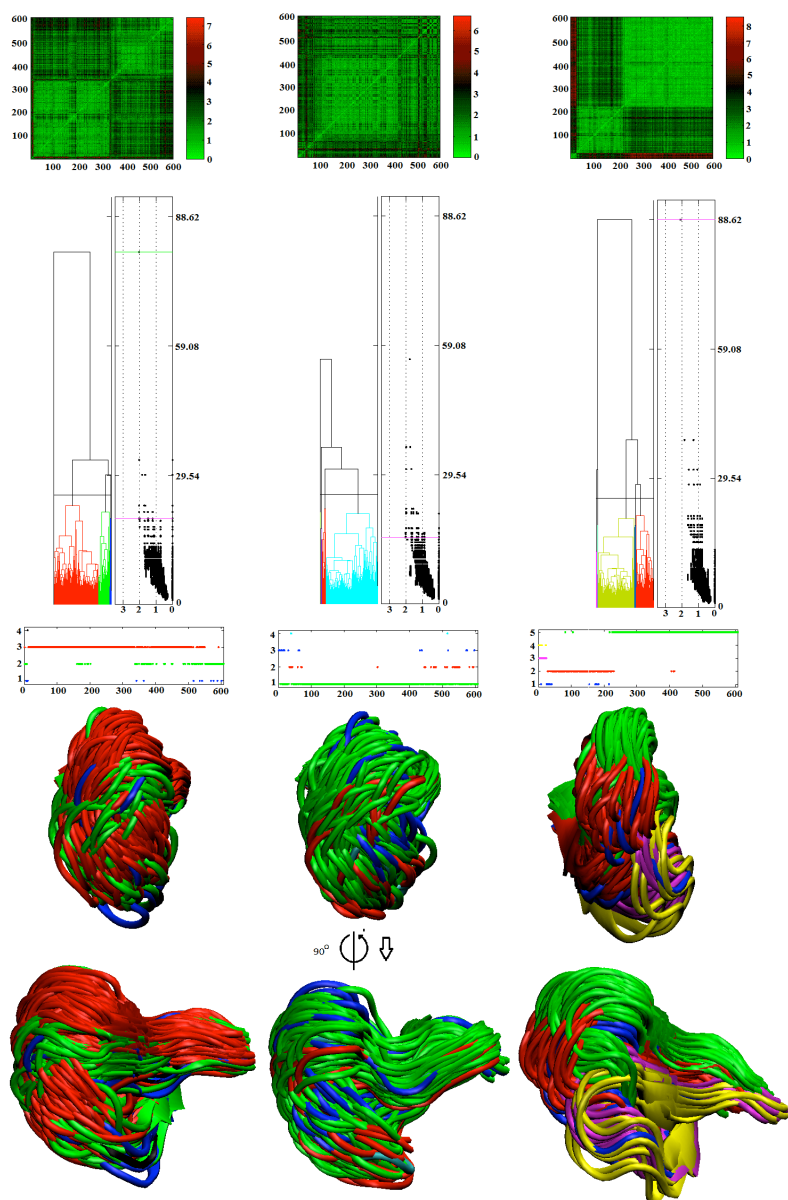


Figure 4.11. Structural variation in switch region three. The columns correspond to α -helical domain ordered from left to right: $G_{\alpha_{\text{GDP}}}$, $G_{\alpha_{\text{GTP.MG}}}$ and $G_{\alpha_{\text{GDP}}}\cdot G_{\beta\gamma}$. The top row corresponds to 2D-RMSDs deviations between frames in each trajectory. The second row shows the dendrograms that group distances between groups of frames. Clusters are defined and colored with a threshold distance of 25. Average distances between frames are plotted against the corresponding average distances in angstroms in the envelope plots next to the dendrograms. As a guide to relate thresholds to angstroms, color coded thresholds are shown for the mean average of frame averages below 3 (red), 2.5 (blue) and 2 (green) Å. Also, for maximum average of frame averages below 3 (black), 2.5 (cyan) and 2 (magenta) Å. The following row shows how the clusters from the dendrograms alternate through the simulation. The structures in the row below show how the clusters separate structures in the simulation from two different points of view.

The three smaller clusters sample at the beginning of the simulation and they explain the strained conformations observed in the 2D-RMSD plot. The sampling of the frames throughout the simulation shows that the two monomers sample mostly one cluster for most of the simulation. The heterotrimer shows the yellow, magenta and blue clusters sampled mostly at the beginning of the simulation, but it is bimodal for the most part, green and red clusters. The maximum RMSDs for the three systems are 7.43 Å, 6.71 Å and 8.55 Å for $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP}.G_{\beta\gamma}$ respectively.

Figure 4.12 show that while the sampling of the frames by the three systems overlap, the blue and yellow cluster, corresponding to the monomers, seem to be closer to each other than to the heterotrimer. The RMSD for the three regions combined range from 0 to 10.71 Å.

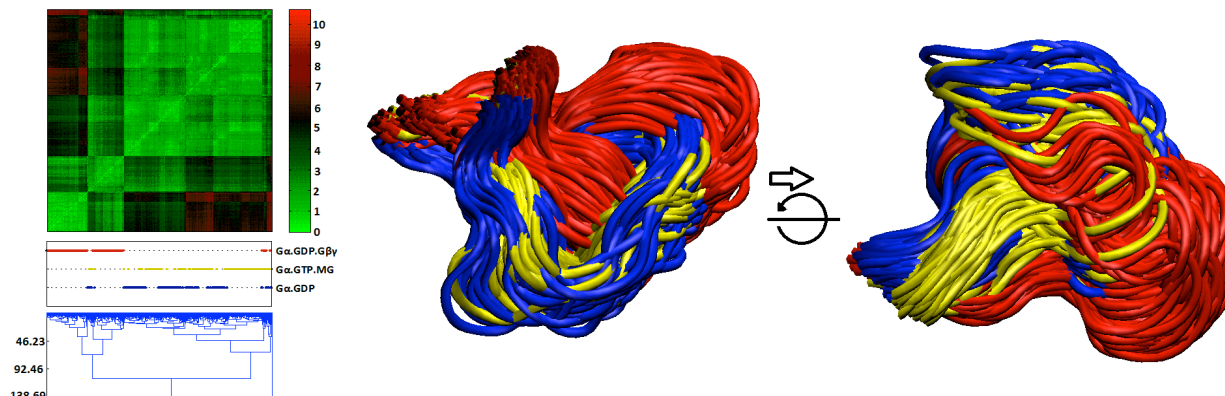


Figure 4.12. Switch region three is mobile in all structures. Switch region three clustering of combined 2D-RMSDs from different trajectories. Left, 2D-RMSDs from Figure 4.11 are used to form a 3 by 3 block matrix for clustering. Clustering shows that the three systems sample clusters that overlap with each other. This region shows four defined clusters within thresholds of 45.95 and 138.69. Right, two angles of switch region three show that while the sampling of the frames by the three systems overlap, the blue and yellow cluster, corresponding to the monomers, seem to be closer to each other than to the heterotrimer.

4.4.2.4. Switch region four

This region is not considered to be as important as the previous regions in crystallographic papers. It is located far from the nucleotide in the three systems, and also far from $\beta\gamma$ in the heterotrimer. Its analysis is relevant because it is the only other major loop that is not close to the N-terminus domain. As a reminder, the N-terminus moves a lot and it is meant to be restraint by the cell membrane; analysis of anything close to it would be hard to interpret from molecular dynamic simulations in solution. Figure 4.13 shows the 2D-RMSD for the three

systems increasing in mobility in the following order: $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$, and $G_{\alpha.GDP.G\beta\gamma}$; this increase in mobility is supported by the ranges of RMSD for the three systems—which in the same order as mention before—are 3.58 Å, 5.88 Å and 7.96 Å.

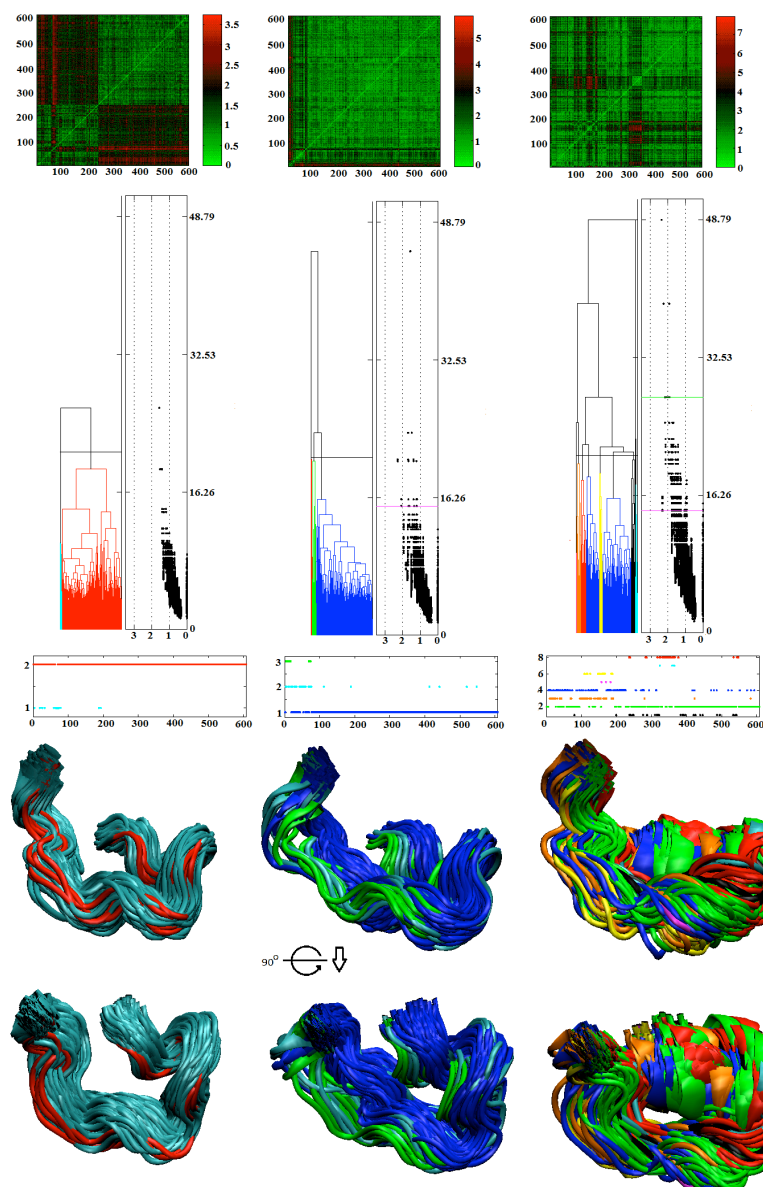


Figure 4.13. Structural variation in switch region four.

The columns correspond to α -helical domain ordered from left to right: $G_{\alpha.GDP}$, $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$. The top row corresponds to 2D-RMSDs deviations between frames in each trajectory. The second row shows the dendrograms that group distances between groups of frames. Clusters are defined and colored with a threshold distance of 21. Average distances between frames are plotted against the corresponding average distances in angstroms in the envelope plots next to the dendrograms. As a guide to relate thresholds to angstroms, color coded thresholds are shown for the mean average of frame averages below 3 (red), 2.5 (blue) and 2 (green) Å. Also, for maximum average of frame averages below 3 (black), 2.5 (cyan) and 2 (magenta) Å. The following row shows how the clusters from the dendrograms alternate through the simulation. The structures in the row below show how the clusters separate structures in the simulation from two different points of view.

Because of the large mobility of the heterotrimer, a threshold of 19 was used to get a number of clusters that is not too large. For the heterotrimer, a threshold higher than 19 would put under one cluster, in green, two clusters that seem far apart. Lowering the threshold more would increase the number of cluster significantly for the heterotrimer without changing the

number of clusters for the two monomers. Too many clusters for one system make their analysis impractical.

With a threshold of 19, only one cluster is obtained in $G_{\alpha,GDP}$, four in $G_{\alpha,GTP.MG}$, and eight for $G_{\alpha,GDP.G\beta\gamma}$. It can be seen that for the $G_{\alpha,GTP.MG}$ the blue cluster separates from the green and red one, with the cyan cluster somewhat as an intermediary. The clusters in the heterotrimer show more disorder, making them harder to characterize. In general, all clusters seem to have the helix region of them more defined in the heterotrimer than in the two monomers.

Figure 4.14 shows frames from the three systems overlapping to some degree, with $G_{\alpha,GTP.MG}$ sampling conformations with the heterotrimer more often, and $G_{\alpha,GDP.G\beta\gamma}$ and $G_{\alpha,GDP}$ having frames belonging to the same clusters. $G_{\alpha,GDP}$ shows less mobility than $G_{\alpha,GTP.MG}$, yellow, and this one in turn with less mobility than $G_{\alpha,GDP.G\beta\gamma}$, red. The RMSD for the three regions combined range from 0 to 7.96 Å.

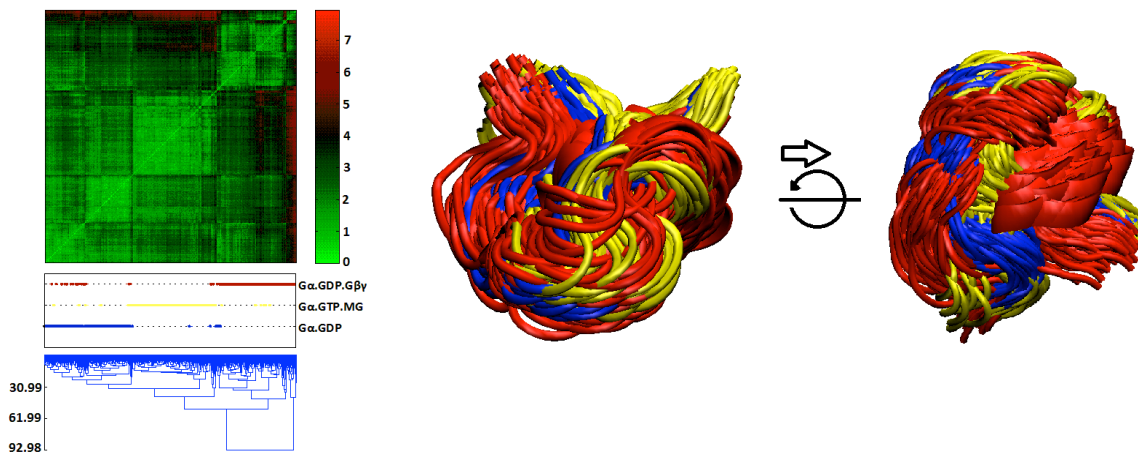


Figure 4.14 Switch region four is variable in all structures. Switch region four clustering of combined 2D-RMSDs from different trajectories. Left, 2D-RMSDs from Figure 4.13 are used to form a 3 by 3 block matrix for clustering. Clustering shows that frames from the three systems overlap to some degree, with $G_{\alpha,GTP.MG}$ sampling conformations with the heterotrimer more often, and $G_{\alpha,GDP.G\beta\gamma}$ and $G_{\alpha,GDP}$ having frames belonging to the same clusters. This region shows four defined clusters within thresholds of 42.80 and 92.98. Right, two angles of switch region four shows $G_{\alpha,GDP}$ with less mobility than $G_{\alpha,GTP.MG}$, yellow, and this one in turn with less mobility than $G_{\alpha,GDP.G\beta\gamma}$, red.

4.4.3 Comparison of B-factors between crystallographic structures and simulations

Figure 4.15 shows the comparison between crystallographic and simulation B-factors. The magnitudes of the B-factors indicate more fluctuation in the simulations than in the

crystallographic structure. A notable difference in both type of B-factors plotted in Figure 4.15 are the regions defined by residues 60 to 178. In the crystal structure, this region does not show significant differences in B-factor values from the rest of the sequence; in the calculated B-factor, this regions show increased B-factors. The bottom region of Figure 4.15 shows a closer view of the B-factors for the GTPase domain and loop regions of interest. The sections of the sequence corresponding to the loop regions show increased mobility.

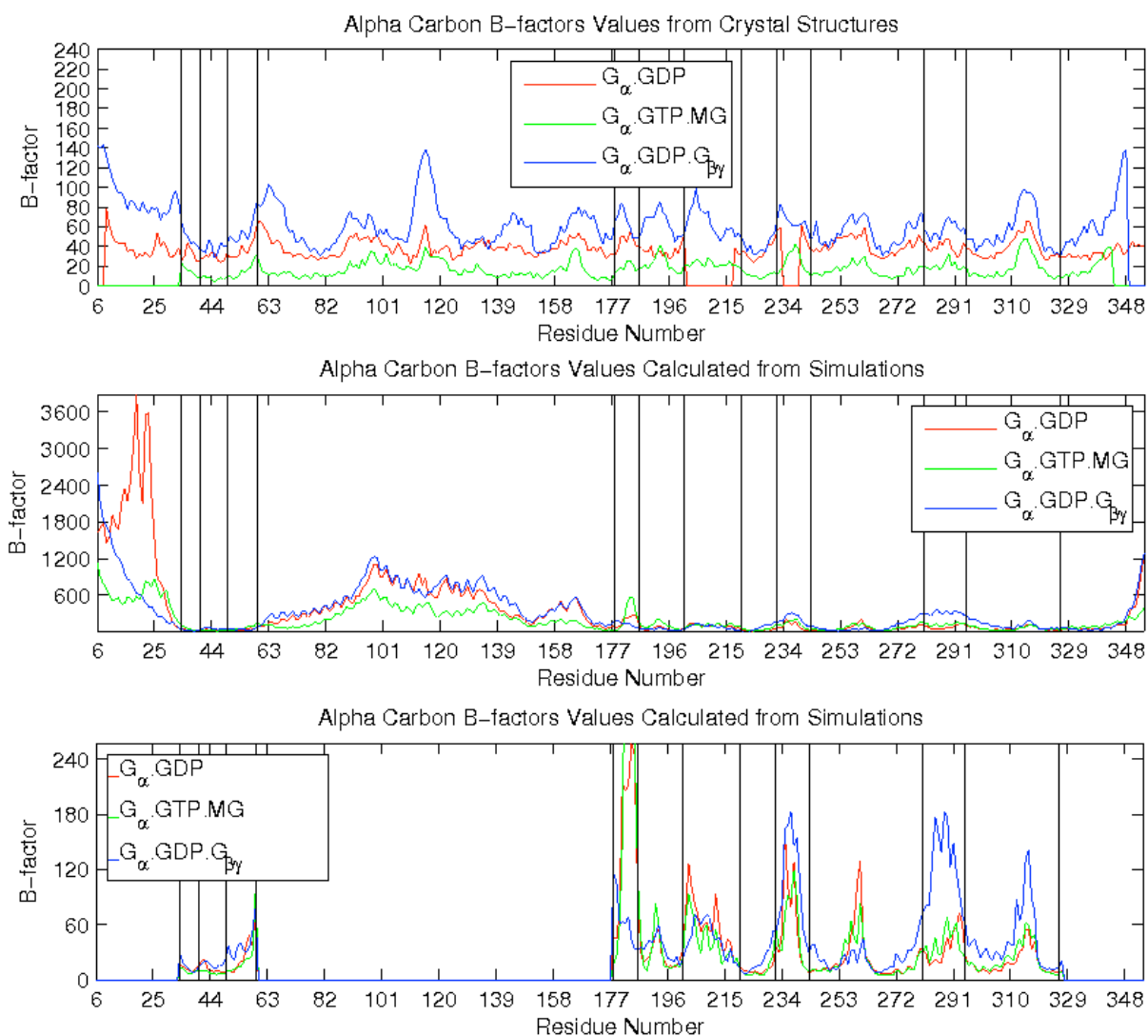


Figure 4.15. Comparison between B-factors obtained from crystallographic structures and simulations for alpha carbons of every residue in the backbone. Top, crystal structure B-factors; middle, B-factors from simulations; bottom, B-factors from simulation are plotted not including N and C-terminus and the residues 60 to 178. The vertical black lines correspond to residue numbers that divide the G_{α} in regions as described in table 4.1. These regions separate sections of interest such as the N and C terminus, loop regions, and the GTPase and helical domain.

4.5. Conclusion

Backbone structure characterization is performed with the relatively rigid GTPase domain as a frame of reference for the three trajectories of the heterotrimeric G-protein in three different states. The motions of regions of interest were characterized in a quantitative way by using RMSD matrices and clustering. The **α helical region** attached to the GTPase domain shows differences in motion for the three regions: with the monomer bound to GTP in a more rigid conformation than the other two systems, and the other two systems populating different extremes of the range of motion observed. The loop regions show differences in mobility that are particular to every system. For **switch region one** the heterotrimer is very rigid, the monomer bound to GTP shows a broader distribution of clusters and the inactive monomer as an intermediate between the two in terms of mobility. In this region, the mobility of $G_{\alpha.GTP.MG}$ is directly associated to loss of coordination between Thr 181 and the Mg^{2+} ion. Switch region one is close to the nucleotide and far from the $\beta\gamma$ subunit in the heterotrimer; this suggest an allosteric stabilization effect of this region cause by the presence of $\beta\gamma$. **Switch region two** is less mobile in the heterotrimer because this region is stabilized by the $\beta\gamma$ subunit. In the monomers, this region has more mobility in $G_{\alpha.GDP}$ than in $G_{\alpha.GTP.MG}$; this is consistent with the fact that the inactive monomer does not have electron diffraction for this region and therefore it moves more. **Switch region three** can be characterized not by the degree of its mobility but by the number of clusters it forms in the three systems; the two monomers sample conformation that are predominantly grouped in one cluster while the heterotrimer samples two distinct clusters. **Switch region four** is less mobile for the inactive monomer, $G_{\alpha.GDP}$, than for the active monomer, $G_{\alpha.GTP.MG}$, and this one in turn is less mobile than the heterotrimer.

The four regions characterized by clustering are restrained on both ends to the GTPase domain. When these regions are compared across systems by clustering conformation from separate trajectories, they show not only differences in general range of conformation, but also in secondary structure. It appears that these loop regions sample well defined conformational basins, and while there is no certainty that this—and all important—basins are fully explored, the ranges of motions, and the restrained in both ends of these regions, make it unlikely that a radical departure from the observed conformation occur in molecular dynamics at equilibrium. The question that remains is how these regions would depart from equilibrium when interacting with other proteins in the cell and the membrane, by aiding transitions in the heterotrimeric G-protein.

If the loop regions undergo radical changes when G-protein-interacting proteins come in contact, then this work would be a reference starting point for further studies of the heterotrimeric G-protein in complex with other proteins.

CHAPTER 5

Energetic analysis of relevant components from extended molecular dynamic simulations of heterotrimeric G-proteins

5.1 Abstract

Extended simulations of a heterotrimeric G-protein in three different states were performed for up to 608 ns. The simulated states of the heterotrimeric G-protein correspond to the inactive $G_{\alpha.GDP}$ monomer, the active $G_{\alpha.GTP.MG}$ and the heterotrimer in the bound state $G_{\alpha.GDP}.G_{\beta\gamma}$. The electrostatic free energies of binding and folding were calculated for all residue components using the generalized-Born (GB) method. From mutation energy ranking of the main components in GB, Poisson–Boltzmann (PB) calculations were performed on components within a GB cutoff ± 1 kcal/mol. The implementation of this filter is important given the large size of the three systems, which makes it very impractical to perform calculations of this simulation length for all the components. An energetic profile of the heterotrimeric G-protein is presented in this chapter from the main components; this energetic analysis allows identifying important residue components, their network of interactions and role in the function of the protein. An energetic analysis serves as a preliminary way to single out important residues that could escape the eye from direct observations of the trajectories. The contributions of this theoretical work can be applied to mutagenesis studies of residues that did not seem important from crystallographic structures.

5.2 Introduction

There are many proteins interacting continuously inside the cell. These proteins are responsible for most of the cell's functioning by associating and dissociating in a concerted way.⁽¹³⁴⁾ The proteins' mechanism for recognizing and interacting with each other is known to work with a high degree of accuracy but is poorly understood at the atomistic level. Electrostatic interactions, both of the solvent and the solute, lie at the heart of this intricate network of protein interactions and their high level of specificity.⁽¹³⁵⁾ Understanding electrostatic interactions between proteins opens the possibility of revealing the mechanisms that cascade into signals that allow the cell to function and to interact with its environment.⁽¹³⁶⁾

It is possible to study these electrostatic interactions from crystallographic structures that have protein-protein or protein-ligand binding interfaces. The information obtained from crystal structures can be expanded with the help of computational methods. The simulations of the heterotrimeric G-protein in its different states provide a time-dependent theoretical model of the G-protein. All atom simulations have been useful in understanding structure and dynamics of proteins.(54, 66-67) It is then expected that with the help of these methods additional insights can be obtained to what crystallographic structures provide.

Electrostatic interactions merit further study from long molecular dynamic simulations for an additional reason. Had it not been for electrostatic interactions, and the corresponding solvent contributions, and interactions between proteins were limited to the ones driven by van der Waals electrostatics only, proteins inside the cell would interact in a more promiscuous way, making it difficult for the cell to function correctly. To better understand the role of electrostatics in protein specificity, calculations of free energies in solution need to be done on the trajectories obtained from simulations. Additionally, a breakdown of energetics can be done to obtain free energy contributions from individual residue components in the protein using the generalized-Born (46, 80-81) and the linearized Poisson–Boltzmann (89-91) methods.

Two types of free energy calculations are utilized to dissect the electrostatic contribution of each component: binding and folding, the latter is also referred as stability. The contribution of every component to the net electrostatic free energies of binding is calculated for protein-nucleotide and protein-protein interactions. This gives a total of four binding interfaces that require separate calculations, two of those for the heterotrimer and one for each monomer. The contribution of every side chain to folding, or stability, of the protein is also done for the three systems. In summary, the total binding and folding electrostatic free energy calculations for the three systems, for each of the 608 frames, adds up to almost five million. The CPU and memory requirements to perform these calculations are difficult to attain for the computationally more expensive method of PB. Each component takes on average 5 minutes to calculate with the Poisson–Boltzmann method, which adds up to 15687.4 days for all components in all systems for every frame using a single CPU. For this reason, it would be impractical to perform all of these calculations knowing that a large number of the components are far from the binding interface, and many non-polar components have little electrostatic contributions. It is then that the less

accurate GB method emerges as a quick alternative to select the most important components based on energies to be refined with PB. $G_{\alpha, \text{GDP}}$ has 1031 components that take in total approximately 30 minutes to calculate using the Generalized-Born method.

The use of GB calculations as a filter of relevant residue components for calculation with PB had been tested before in a previous work that is also presented in Chapter 3.(137) Generalized-Born component analysis models have also been applied by Karplus and co-workers to a small peptide, but their results were not compared to similar calculations with the PB method,(138) and neither were the results of Zoete and Michielin.(139) On the other hand, the exclusive use of PB has been applied to the study of the GCN4 leucine zipper (64 residues) by Hendsch and Tidor,(33) in rational redesign of calmodulin (169 residues), by Green, Tidor and Jasanoff, and to optimize an inhibitor of HIV entry into the cell (210 residues).(34) These applications of PB were done on systems that are relatively small compared to the heterotrimeric G-protein and its three states. They did not require a filter because the number of calculations was not as demanding.

The degree to which generalized-Born approximates Poisson–Boltzmann for each residue component was presented in the work by Carrascal and Green.(137) That work is also presented in Chapter 3 of this thesis. GB and PB electrostatic binding free energies were calculated for all the components for 100 ns simulation time, for all frames taken at 1 ns intervals. The results of these calculations quantified how much GB approximates PB; while GB is not a very accurate substitute for PB calculations, it was concluded that the use of GB is justified as a filter for selecting the most important components to be calculated with PB—giving a good tradeoff between accuracy and speed for large systems.

Mutation energies are calculated to assess the contributions of individual components to the electrostatic free energies of binding and folding. This type of energy is the sum of direct, indirect and desolvation energies, as described in Equation 5.1. The coefficients i,j correspond to indexes between every pair of components in the system. Self electrostatic free energies are calculated for each component only once, ΔG_i^{solv} , and the pair-wise electrostatic free energies are calculated for every permutation of components i,j .

$$\Delta G_i^{mut} = \Delta G_i^{solv} + \sum_j \Delta G_{ij}^{dir} + \sum_j \Delta G_{ij}^{indir} \quad (5.1)$$

The terms in these equations are solvation, direct and indirect contributions. Solvation energy corresponds to the self electrostatic free energy difference in contribution from each residue component in the bound and the unbound states, or the unfolded to the folded state. For each pair of components, direct energies correspond to solvent-screened Coulombic pair-wise interactions between component i and all of the components in the binding partner. Indirect energy corresponds to solvent-screened Coulombic pair-wise interactions between component i and every other component in the same molecule. Mutation energy is the theoretical counterpart of alanine scanning mutagenesis. This energy corresponds to the theoretical process of turning on the charges of a component, and it is a measure of the component importance to binding or folding.

Equation 5.1 is applicable to both folding and binding. In folding, the equation corresponds to the difference in energy from the unfolded side chain alone and the side chain attached to the protein and folded, surrounded by solvent, the remaining of the protein or both, depending on the side-chain location. Binding is calculated from the energetic difference of each residue component in the presence and absence of the binding molecule.

Individual components' average free energies of binding and folding provide a measure of their importance to each system. The importance of each component to transitions of heterotrimeric G-protein states may reveal changes important to signal transduction from a thermodynamic state function; that is, a function in which changes are studied for end states of a protein and independent of the energy necessary for the transition. These differences of states in free energies of binding between the systems describe what is energetically different, or how much a component's energy changes.

Thorough calculations of electrostatic binding and folding free energies are important in understanding signal transduction at the atomic level. The passing of signal through the cell membrane requires a relaying heterotrimeric G-protein that in turn undergoes specific intra-protein mechanisms between its electrostatic components. Passing of a signal within a protein is as important as passing a signal between proteins; the following work sheds light on both intra and inter protein electrostatic signal relay by studying changes in the α -alpha subunits from three states, and how components interact between the α and $\beta\gamma$ subunits in the heterotrimer. This type

of analysis gives a higher certainty that no important interactions were missed, and it gives a succinct way of organizing electrostatic interactions in a filtered profile by energetic relevance; this could be used as a reference for possible characterizations of other members of the G-protein family that have slightly different sequences or mutations; thus, extending understanding in signal transduction from theoretical models.

5.3 Methods

5.3.1 Poisson–Boltzmann calculations

All solutions of the linearized Poisson–Boltzmann equation were obtained using a multigrid finite-difference solver distributed with the ICE (Integrated Continuum Electrostatics) package.⁽¹¹⁴⁾ The atomic radii used were those optimized by Roux and co-workers for use with continuum electrostatic models;⁽¹¹⁶⁻¹¹⁷⁾ partial atomic charges were taken from the PARAM22/27 parameter set.⁽¹¹²⁻¹¹³⁾ For consistency with the all-atom molecular dynamics simulations, an internal dielectric constant of 1 was used, along with an external dielectric constant of 80. Focusing boundary conditions were used: an initial calculation was done with the largest dimension of the system occupying 23% of one side of the grid, and Debye–Hückel boundary conditions were used; a second calculation was then done with the system occupying 92% of the grid, and using the potentials from the previous calculation at the boundary. For individual components, but not for net binding energies, an additional overfocused calculation was done with the molecule occupying 184% of the grid (centered on the component of interest). Net binding energies were computed with a 257^3 unit cubic grid, and individual components with a 129^3 unit grid; due to the overfocusing used for components, the grid spacing was identical in both of these cases.

5.3.2 Generalized-Born calculations

Generalized-Born calculations were done using the GBSW module ⁽¹⁰⁵⁾ of a version of the CHARMM computer program modified to output the effective Born radii. As for the PB calculations, the atomic radii used were those optimized by Roux and co-workers ⁽¹¹⁶⁻¹¹⁷⁾ Binding was considered as the difference between the bound state and a state in which one component was rigidly translated 500 Å. The scaling coefficients were set to standard values, of

$a_0 = 1.2045$ and $a_l = 0.1866$, the molecular surface was used, and a smoothing length of 0.2 \AA was applied; no cutoffs were used.

5.3.3 Tabulation and ranking of components within the cutoff

All components in the three systems were initially calculated with generalized-Born model. A cut off of 1 kcal/mol was used to select the top components from GB, for binding and folding electrostatic free energy. These components were then calculated using Poisson–Boltzmann component analysis. Only the top 50 components in folding for the trimer were recalculated with PB. The ranking of components in these tables is done in terms of the absolute value of their energies. In this way, both favorable negative energies and unfavorable positive ones are considered relevant for their magnitude.

5.4 Results and Discussion

The decomposition of binding and folding free energies was performed on the heterotrimeric G-protein in three different states on structures obtained from molecular dynamic simulations. For each one of the structures, the contribution of every residue to binding and folding was tabulated and ranked for analysis using GB initially. PB calculations were performed on a set of components selected from top ranked GB results. The selection criteria applied to GB results for refinement with PB includes components with energies larger than 1 kcal/mol for unfavorable interactions, and components with energies lower than -1 kcal/mol for favorable interactions. The total number of PB and GB components considered for the analysis, as well as the number of backbone and side chain components, is presented in Table 5.1 for folding and binding. For binding, all of the components within the GB cutoff were calculated with PB. For folding of the heterotrimer, a lot of components fall within the cutoff, and it is not practical to calculate them all with PB. For this reason only the top 50 components for each system were calculated.

5.4.1 Important components for binding

5.4.1.1. Nucleotide-protein binding energy

Mutation energies of binding of side chain components are shown in Figure 5.1 on the left. This figure can be seen as an electrostatic profile that shows the nucleotide's points of

contact to G_{α} . These points of contact across the heterotrimeric G-protein give an initial indication of how electrostatics within the protein varies as signals are passed through them. Some of the components that are significant in binding for the three systems have conserved interactions that do not change from one state of the G-protein to another. In order to determine residues with relevant changes, the plots in Figure 5.1 A, B and C, on the left, were subtracted from one another to give side chain differences in mutation energies from one state of the heterotrimeric G-protein to the other; these are shown in Figure 5.1 on the right. In other words, the figure corresponds to changes in energy before and after nucleotide exchange and the consequent trimer dissociation and association.

Table 5.1. Number of components calculated with GB and PB.

System	Binding						Folding ^a			
	GB			PB			GB			PB
	total	Backbone	Side chains	total	Backbone	Side chains	total	Backbone	Side chains	Side chains
$G_{\alpha.GDP}$	1030	698	332	47	31	16	785	520	265	50
$G_{\alpha.GTP.MG}$	1030	698	332	45	29	16	788	516	272	46
$G_{\alpha\beta\gamma}$ to GDP	2185	1486	699	39	23	16	1639	1108	531	46
$G_{\alpha.GDP}$ to $G\beta\gamma$	2185	1486	699	53	44	9				

^aGB folding gave a large number of components for the ± 1 kcal/mol cutoff that include side chain and backbone components. PB folding is calculated for side chains selected from the top 50 GB folding components; this number of components corresponds to a cutoff energy of 6.86 kcal/mol for $G_{\alpha.GDP}$, 6.22 kcal/mol for $G_{\alpha.GTP.MG}$ and 8.67 for $G_{\alpha.GDP.G\beta\gamma}$. PB folding gave 46 components for $G_{\alpha\beta\gamma}$ to GDP and $G_{\alpha.GDP}$ to $G\beta\gamma$ because some components that are included in the 1 kcal/mol GB cutoff do not make the cutoff in PB.

The association of $\beta\gamma$ causes an unfavorable change in many components that are bound to the nucleotide because they go from interactive primarily with the nucleotide, to having $\beta\gamma$ competing in their interactions. In Figure 5.1 for example, two components gain in mutation energy to go into a less favorable energetic state upon binding of the heterotrimer, Glu 43 and Asp 272, both of which simply go from favorable state to a less favorable state, and giving a peak with an unfavorable energy in the plot of Figure 5.1 F.

Components selected outside the ± 1 kcal/mol energy range can be grouped in five major discrete regions along the primary structure sequence. These regions are separated in the protein

sequence, but come close to the nucleotide making significant interactions. In Figure 5.1, the nucleotide interacts with side chain residues that range from closer in sequence to the N-termini to closer in sequence to the C-termini. The five regions that show significant binding PB, GB energies are analyzed in groups and more detail in the remaining of this section.

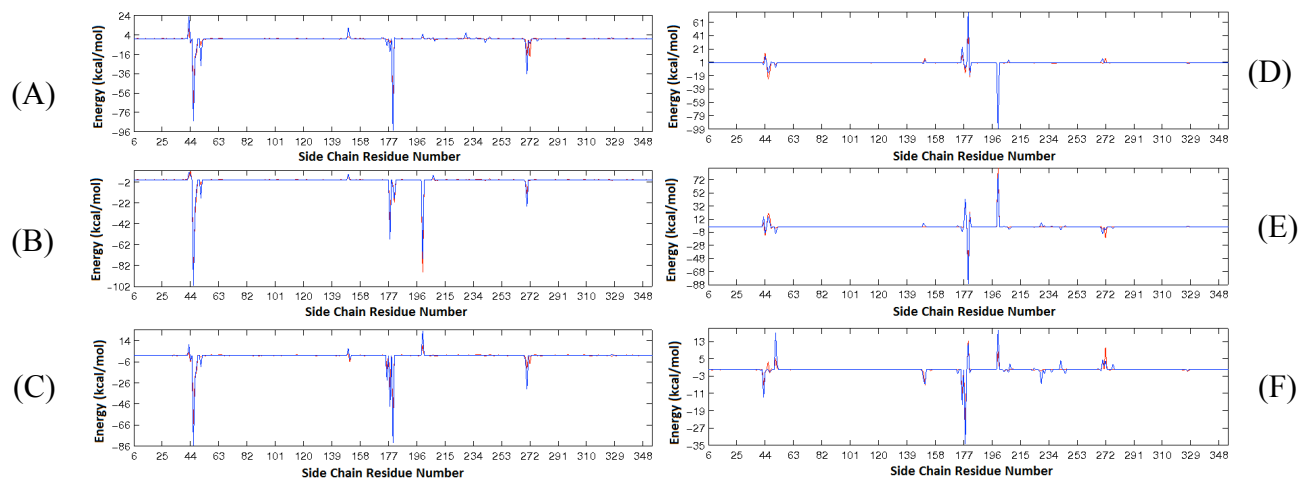


Figure 5.1. Contribution to binding free energies. PB(blue) and GB(red) side chain mutation energies for binding of G_{α} to the nucleotide for all GB components(red) and selected PB components(blue). Left, (A) $G_{\alpha}.GDP$. (B) $G_{\alpha}.GTP.MG$ (C) $G_{\alpha}.GDP.G_{\beta\gamma}$. Right, (D) $\Delta G(G_{\alpha}.GTP.MG) - \Delta G(G_{\alpha}.GDP.G_{\beta\gamma})$, the G-protein goes from the inactive heterotrimer to the active monomer. (E) $\Delta G(G_{\alpha}.GDP) - \Delta G(G_{\alpha}.GTP.MG)$, the G-protein hydrolyses the third phosphate group in the nucleotide and goes from the active to the inactive state. (F) $\Delta G(G_{\alpha}.GDP.G_{\beta\gamma}) - \Delta G(G_{\alpha}.GDP)$, the G-protein goes from the monomeric state to the association of $\beta\gamma$ to form the heterotrimer again.

5.4.1.1.1. Nucleotide-protein binding group one

Side chains Glu 43, Ser 44, Lys 46, Ser 47, Thr48, Lys 51 make interactions with the nucleotide that make the cutoff energy in the three systems. Gln 52 only makes the cutoff energy for the inactive monomer, $G_{\alpha}.GDP$. These side chains have chemical groups that tend to facilitate hydrogen bonding with the phosphate groups of the nucleotide. The backbone superposition in Figure 5.2 shows that the parts of this sub domain that hold the nucleotide in place are relatively rigid in the three systems. The third phosphate group displaces the backbone slightly more for the active monomer, but this backbone section is very rigid relative to the loop regions one to three, so not much can be inferred from this differences.

There are few noticeable structural differences of side chain conformation in the three systems for this region, and the differences are mostly in energetics listed in Table 5.2. Glu 43

has unfavorable binding energies in the three systems and point away from the phosphate groups. The side-chain and amino group of Lys 46 interact favorably with the last phosphate group of the nucleotide in both inactive states of the G-protein. In the active state, however, the amino group of Lys 46 interacts with the second to last phosphate group. Ser 47 interacts favorably through a hydrogen bond with the nucleotide in the inactive states, but in the active state, the side chain oxygen atom interacts with the magnesium ion, making its interaction with the nucleotide unfavorable by breaking the hydrogen bond and forcing the oxygen atoms to get closer. Thr 48 interacts in a similar way in the three states of the heterotrimeric G-protein forming a conserved hydrogen bond with the same phosphate group in the nucleotide.

Table 5.2. PB nucleotide mutation binding energies for binding group one.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Glu 43	23.32	7.45	10.29	-2.84	15.87	-13.03
Ser 44	-4.10	5.06	-5.02	10.08	-9.16	-0.92
Lys 46	-85.14	-101.05	-85.30	-15.75	15.9	-0.16
Ser 47	-22.28	-29.97	-23.54	-6.43	7.69	1.26
Thr 48	-11.17	-9.51	-11.64	2.13	-1.66	0.47
Lys 51	-28.02	-17.69	-10.88	-6.81	-10.33	-17.14
Gln 52	-2.96	-0.29*	-0.01*	0.28*	-2.67*	2.95*

^a(A) $G_{\alpha,GDP}$, (B) $G_{\alpha,GTP,MG}$ and (C) $G_{\alpha,GDP,G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha,GTP,MG}) - \Delta G(G_{\alpha,GDP,G\beta\gamma})$. (E) $\Delta G(G_{\alpha,GDP}) - \Delta G(G_{\alpha,GTP,MG})$. (F) $\Delta G(G_{\alpha,GDP,G\beta\gamma}) - \Delta G(G_{\alpha,GDP})$. * Energies calculated with GB. All units in kcal/mol

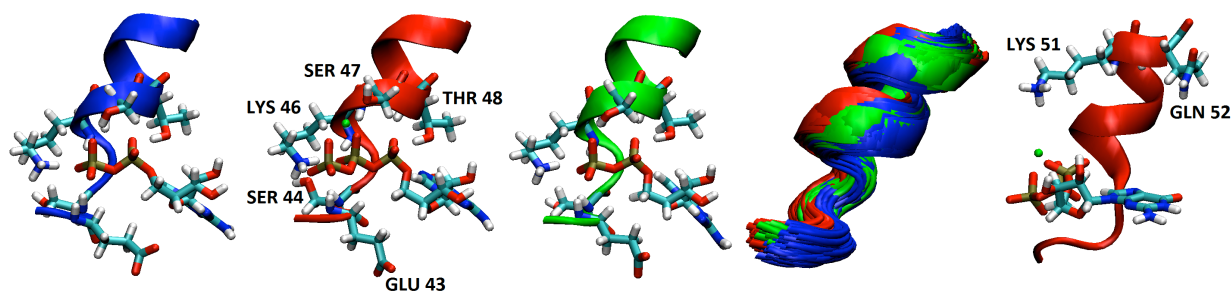


Figure 5.2. Binding group one interactions. From left to right, Glu 43, Ser 44, Lys 46, Ser 47 and Thr 48 side chains in $G_{\alpha,GDP}$, $G_{\alpha,GTP,MG}$, $G_{\alpha,GDP,G\beta\gamma}$ respectively, followed by superimposed back bone conformations for the three systems, and $G_{\alpha,GTP,MG}$ showing only Lys 51 and Gln 52.

5.4.1.1.2. Nucleotide-protein binding group two

This group is formed only by the side chain of Asp 150 in the three systems. In the heterotrimer, the side chain of Ser 151 also makes the cutoff. For these two residues, only the carbonyl group of Asp 150 is included in the cutoff with favorable interactions for $G_{\alpha,GDP}$ and

$G_{\alpha.GDP}.G_{\beta\gamma}$, as well as the amino group of Ser 151 for the heterotrimer. Arg 176 also interacts also interacts favorable in $G_{\alpha.GDP}$ and $G_{\alpha.GDP}.G_{\beta\gamma}$.

In this group, Asp 150 interacts unfavorably in the three states because it is close to the oxygen atom in the ribose group of the nucleotide. In the two monomers, Ser 151 makes few and sporadic interactions with the nucleotide. Ser 151 interacts with the nucleotide only in the heterotrimeric structure where it gets between the backbone of Arg 176 and a hydrogen in the nucleotide. This interaction is conserved during the whole simulation as seen in Figure 5.3.

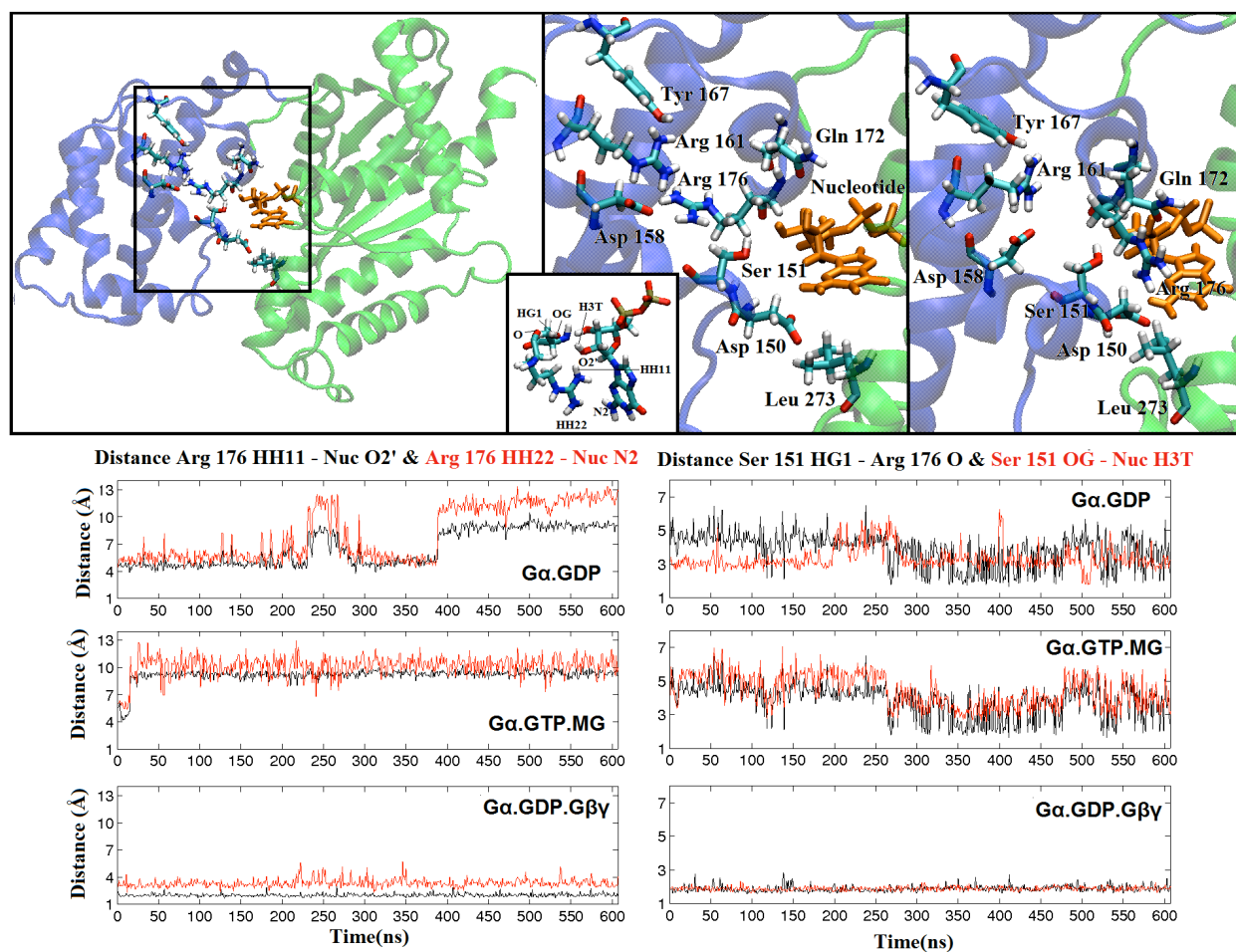


Figure 5.3. Interactions of binding group two. Top left, location of binding group 2 in the α sub-unit. Top center, Arg 176 interacting with residues in the α helical domain and away from the nucleotide. Top right, Arg 176 interacting with the nucleotide. Top inset, labeling of atoms in Arg 176, Ser 151 that interact with the nucleotide. Bottom left, distances from Ser 151 HG1 to Arg 176 O in black, and from Ser OG to Nucleotide H3T in red. Bottom right, distances from Arg 176 HH11 to Nucleotide 176 O2' in black, and from Arg 176 HH22 to Nucleotide N2 in red.

Table 5.3. PB nucleotide mutation binding energies for binding group two.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Asp 150	11.29	5.57	6.81	-1.24	5.72	-4.48
Ser 151	0.95*	0.54*	-4.37	4.37	0.41*	-5.32*
Arg 176	-7.51	-0.51*	-23.87	23.87	23.36	-16.36
Crb 150	-1.90	-0.55*	-2.82	2.27	-1.35*	-0.92*
Amn 151	0.28*	0.10*	-2.15	2.25	0.18*	-2.43*

^a(A) $G_{\alpha,GDP}$, (B) $G_{\alpha,GTP.MG}$ and (C) $G_{\alpha,GDP}.G_{\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha,GTP.MG}) - \Delta G(G_{\alpha,GDP}.G_{\beta\gamma})$. (E) $\Delta G(G_{\alpha,GDP}) - \Delta G(G_{\alpha,GTP.MG})$. (F) $\Delta G(G_{\alpha,GDP}.G_{\beta\gamma}) - \Delta G(G_{\alpha,GDP})$. * Energies calculated with GB. Crb stands for carbonyl group, and Amn for amino group. All units in kcal/mol

Figure 5.3 also shows the difference in interaction of Arg 176. This residue interacts with the nucleotide in the inactive state, $G_{\alpha,GDP}$, for the first 230 ns of the simulation, it then goes to interact with a group of residues in the α -helical domain temporarily, Asp 158, Arg 161, Tyr 167, Gln 172 and Leu 273, before it re-associates and dissociates again. In the active monomer, $G_{\alpha,GTP.MG}$, Arg 176 is also bound to the nucleotide at the beginning of the simulation, but it goes to interact with the group of residues in the α -helical domain only 30 ns into the simulation. In the heterotrimer, $G_{\alpha,GDP}.G_{\beta\gamma}$, Arg 176 never interacts with the group of residues.

5.4.1.1.3. Nucleotide-protein binding group three

This group is formed by various side-chain and backbone groups from 175 to 185, from which only the most significant ones are considered in Table 5.4. Only the side chains of Arg 178 and Lys 180 have significant favorable energies in the three systems. Side-chain Thr 181 is only observed to have significant energies in the two monomer states of the G-protein. The heterotrimer does not have any backbone components within the 1 kcal/mol cutoff, and the inactive monomer only includes carb 179 and carb 180. The active monomer has many backbone groups included by the cutoff likely due to the additional phosphate group.

The structural differences of residues in this group are relevant because they are part of a loop region that connects the α -helical domain and the GTPase domain. Figure 5.4 shows the location of Arg 178, Lys 180 and Thr 181 and the distance from the heavy atoms in their side chains to the second phosphate group of the nucleotide, which is present in the three systems. The schematic of Figure 5.4 is from $G_{\alpha,GTP.MG}$ and shows the coordination of the Mg^{2+} ion by

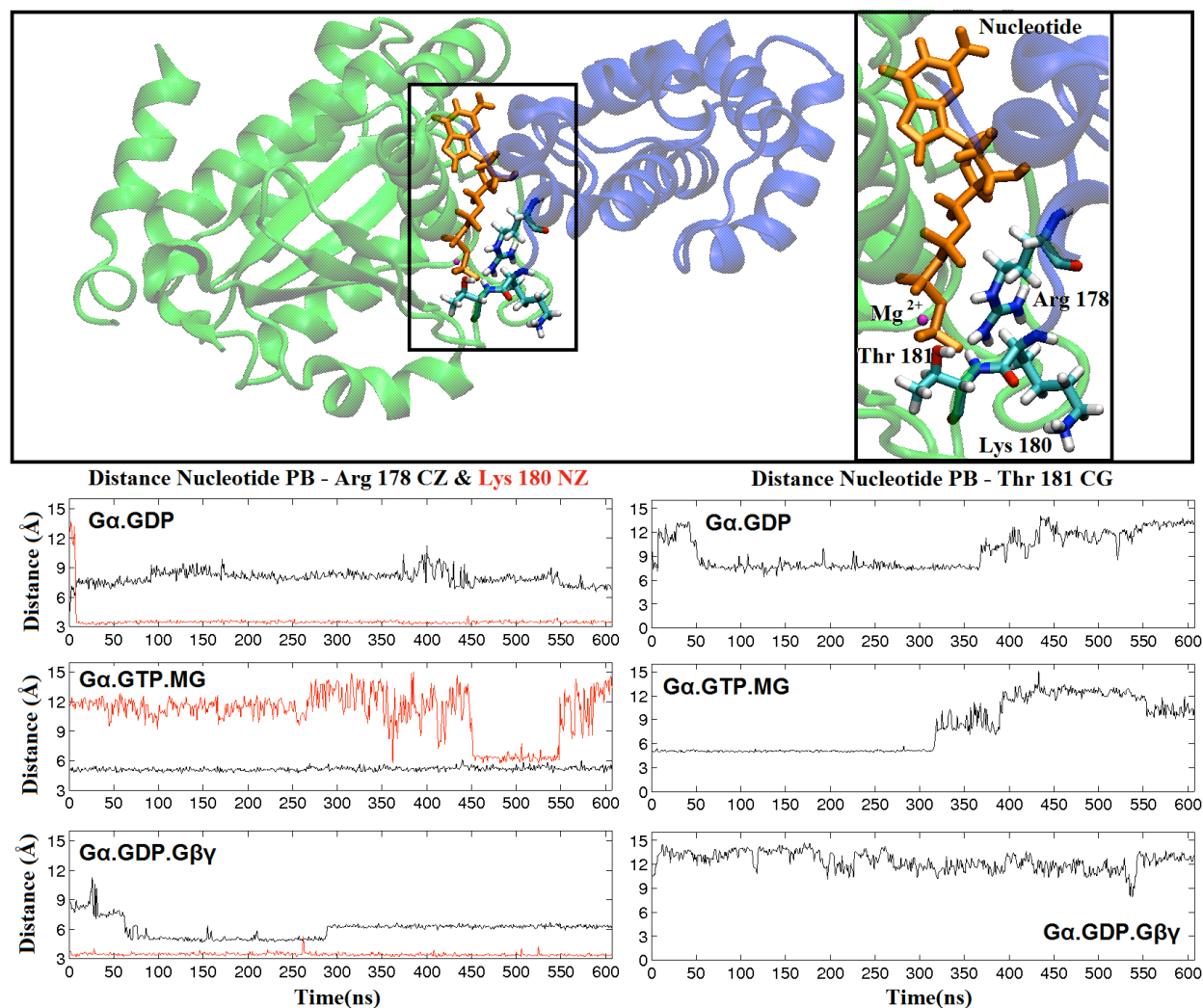


Figure 5.4. Interactions of binding group three. Top left, location of binding group 3 in the α sub-unit with the GTPase domain in green and the α helical domain in blue for $G_{\alpha.GTP.MG}$. Top right, zoom in of Arg 178, Lys 180, Thr 181, the magnesium ion in magenta and the nucleotide in orange. Bottom left, distances from the PB atom in the nucleotide to Arg 178 CZ in black and Lys 180 NZ in red. Bottom right, distances from the PB atom in the nucleotide to Thr 181 CG.

the last two phosphate groups of the nucleotide and Thr 181. The coordination of Thr 181 with Mg^{2+} ion only occurs for a little over the first 300 nanoseconds of the simulations before dissociating for the remaining 308 ns; this dissociation is the reason for the increased mobility of the loop region to which this residue belongs. In the other two states of the G-protein, Thr 181 remains dissociated all the time.

Lys 180 starts far from the nucleotide in the inactive monomer, $G_{\alpha.GDP}$, and gets closer to the nucleotide after 7 nanoseconds into the simulation and never dissociates again. Lys 180 is bound to the nucleotide throughout the whole simulation in the heterotrimeric state of the G-

protein. For the active state, $G_{\alpha.GTP.MG}$, Lys 180 is mostly dissociated, and it only interacts with the nucleotide between the 451th and the 549th nanosecond of the simulation. Lys 180 is always bound to the nucleotide in the heterotrimeric state.

Table 5.4. PB nucleotide mutation binding energies for binding group three.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Arg 178	-13.93	-57.01	-48.49	-8.52	43.08	-34.56
Lys 180	-95.09	-7.78	-82.97	75.19	-90.31	12.12
Thr 181	1.70	-18.50	0.12*	-18.62*	20.20	-1.58*
Crb 175	0.13	-2.69	0.01*	-2.70*	2.82	-0.12*
Crb 176	-1.20	-2.24	0.01*	-2.25*	1.04	1.21*
Crb 178	-0.41*	-3.47	0.01*	-3.4*	3.06*	0.42*
Crb 180	-2.30	-3.71	0.05*	-3.76*	1.41	2.35*
Crb 181	-0.43*	1.89	-0.06*	1.95*	-2.32*	0.37*
Amn 181	-0.22*	-3.27	0.00*	-3.27*	3.05*	0.22*

^a(A) $G_{\alpha.GDP}$, (B) $G_{\alpha.GTP.MG}$ and (C) $G_{\alpha.GDP.G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP.G\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP.G\beta\gamma}) - \Delta G(G_{\alpha.GDP})$. * Energies calculated with GB. Crb stands for carbonyl group, and Amn for amino group. All units in kcal/mol.

Arg 178 is close to Lys 180 and they compete for binding to the nucleotide in the two monomers. In $G_{\alpha.GDP}$, Arg 178 dissociates from the nucleotide at about the same time Lys 180 binds to it. The opposite occurs in $G_{\alpha.GTP.MG}$, Arg 178 remains bound to the nucleotide even when Lys 180 is also bound to the nucleotide between the 451th and the 549th nanosecond. In the heterotrimer, the two appear to remain close to the nucleotide without excluding each other. The distance of Arg 178 to the nucleotide is a little higher than Lys 180, but this is in part due to the larger size of the arginine side chain compared to lysine and the effect of measuring interactions from heavy atoms distances, which makes the distance easier to read when there are side chain flips or rotations.

5.4.1.1.4 Nucleotide-protein binding group four

In the three systems side chain Asp 200 make the cutoff. Only the active monomer, $G_{\alpha.GDP}$, includes any other components, the side chain of Glu 207, the carbonyl groups of 202 and 203 and the amino group 203, shown in Table 5.5.

The location of Asp 200 and Glu 207 is in an intermediate location between the nucleotide and $\beta\gamma$ subunit in the heterotrimer. Asp 200 is closer to the nucleotide while Glu 207

to protein-protein interface as seen in Figure 5.8. In $G_{\alpha,GDP}$, Asp 200 is found on the phosphate side of the nucleotide and it interacts with an unfavorable energy of 4.92 kcal/mol. Glu 207 does not have an interaction energy above the 1kcal/mol cutoff. In $G_{\alpha,GDP}.G_{\beta\gamma}$, Asp 200 also points towards the nucleotide and it is closer to it, giving an even higher unfavorable interacting energy of 23.11 kcal/mol. In both the inactive monomer and heterotrimer, Asp 200 is kept pointing to the nucleotide despite its unfavorable energies by residues surrounding it; this residue is also exposed to the solvent.

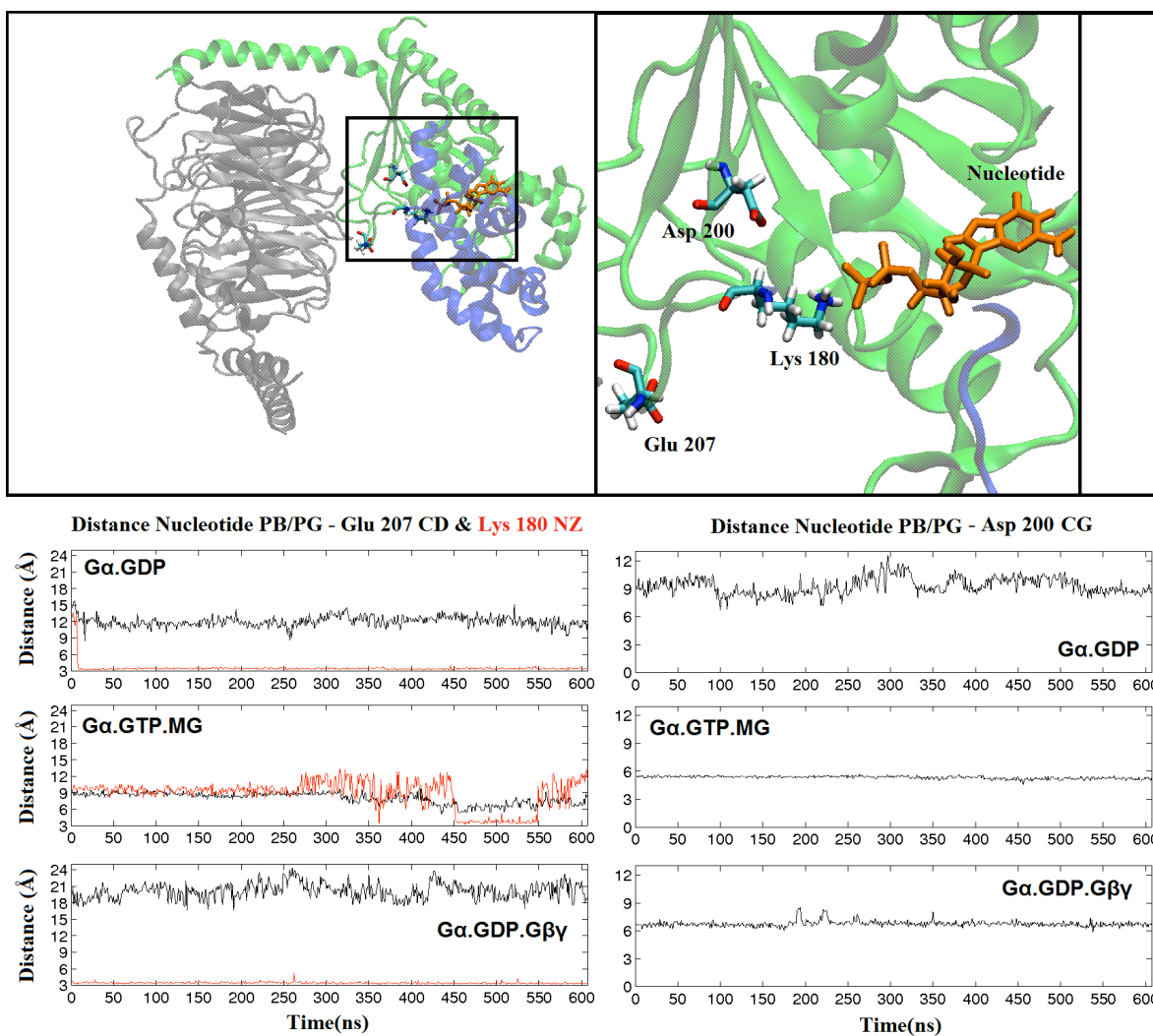


Figure 5.5. Interactions of binding group four. Top left, location of binding group 4 in the α sub-unit with the GTPase domain in green, the α helical domain in blue and $\beta\gamma$ in gray for $G_{\alpha,GDP}.G_{\beta\gamma}$. Top right, zoom in of Asp 200, Glu 207 and Lys 180. Bottom left, distances from the last phosphorous atom (PB/GB depending on the system) to Glu 207 CD and Lys 180 NZ. Bottom right, distances from the last phosphorous atom in the nucleotide to Asp 200 CG.

Table 5.5: PB nucleotide mutation binding energies for binding group four.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Asp 200	4.92	-75.53	23.11	-98.64	80.45	18.19
Glu 207	0.96*	4.40	3.49	0.91*	-3.44*	2.53*
Crb 202	-1.14*	-3.99	-0.01	-3.98*	2.85*	1.13*
Crb 203	-0.73*	-2.37	-0.27	-2.1*	1.64*	0.46*
Amn 203	0.16*	-2.91	0.27	-3.18*	3.07	0.11*

^a(A) $G_{\alpha,GDP}$, (B) $G_{\alpha,GTP,MG}$ and (C) $G_{\alpha,GDP,G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha,GTP,MG}) - \Delta G(G_{\alpha,GDP,G\beta\gamma})$. (E) $\Delta G(G_{\alpha,GDP}) - \Delta G(G_{\alpha,GTP,MG})$. (F) $\Delta G(G_{\alpha,GDP,G\beta\gamma}) - \Delta G(G_{\alpha,GDP})$. Crb stands for carbonyl group, and Amn for amino group. Crb stands for carbonyl group, and Amn for amino group. All units in kcal/mol

In $G_{\alpha,GTP,MG}$, Asp 200 makes a favorable interaction with the Mg^{2+} ion. Glu 207 is involved in the interaction of Lys 180 with nucleotide. Lys 180 is in the solvent until it makes a favorable interaction with Glu 206, shortly after, Lys 180 moves on to interact with the nucleotide. This interaction lasts between 452 and 548 nanoseconds in the simulation; after that Lys 180 interacts intermittently with Glu 207.

5.4.1.1.5. Nucleotide-protein binding group five

Side chains Asn 269, Lys 270, Asp 272 and the 269 carbonyl group are all included in the cutoff for the three systems. Only $G_{\alpha,GDP}$ has two additional components in this group, the carbonyl and amine group of residue number 270. This group of components is located in the side of the α subunit opposite to the $\beta\gamma$ subunits as shown in Figure 5.6. Because of its location, this group of residues could play an important role in nucleotide exchange.

In the three states, Asn 269 makes a favorable interaction with the guanine group of the the nucleotide although in the single digits. The preferred interaction takes place between the hydrogen atom in the carboximide group in asparagine and one of the nitrogen atoms in the nucleotide. The side chain of the asparagine never flips, but it sporadically distances from the nucleotide for a few nanoseconds throughout the simulation in the three systems.

Lys 270 is extended along the nucleotide in the three states of the G-protein, with the hydrophobic part of the side chain along the guanine ring and the polar part in close proximity to an oxygen atom, O4', in the ribose ring of the nucleotide.

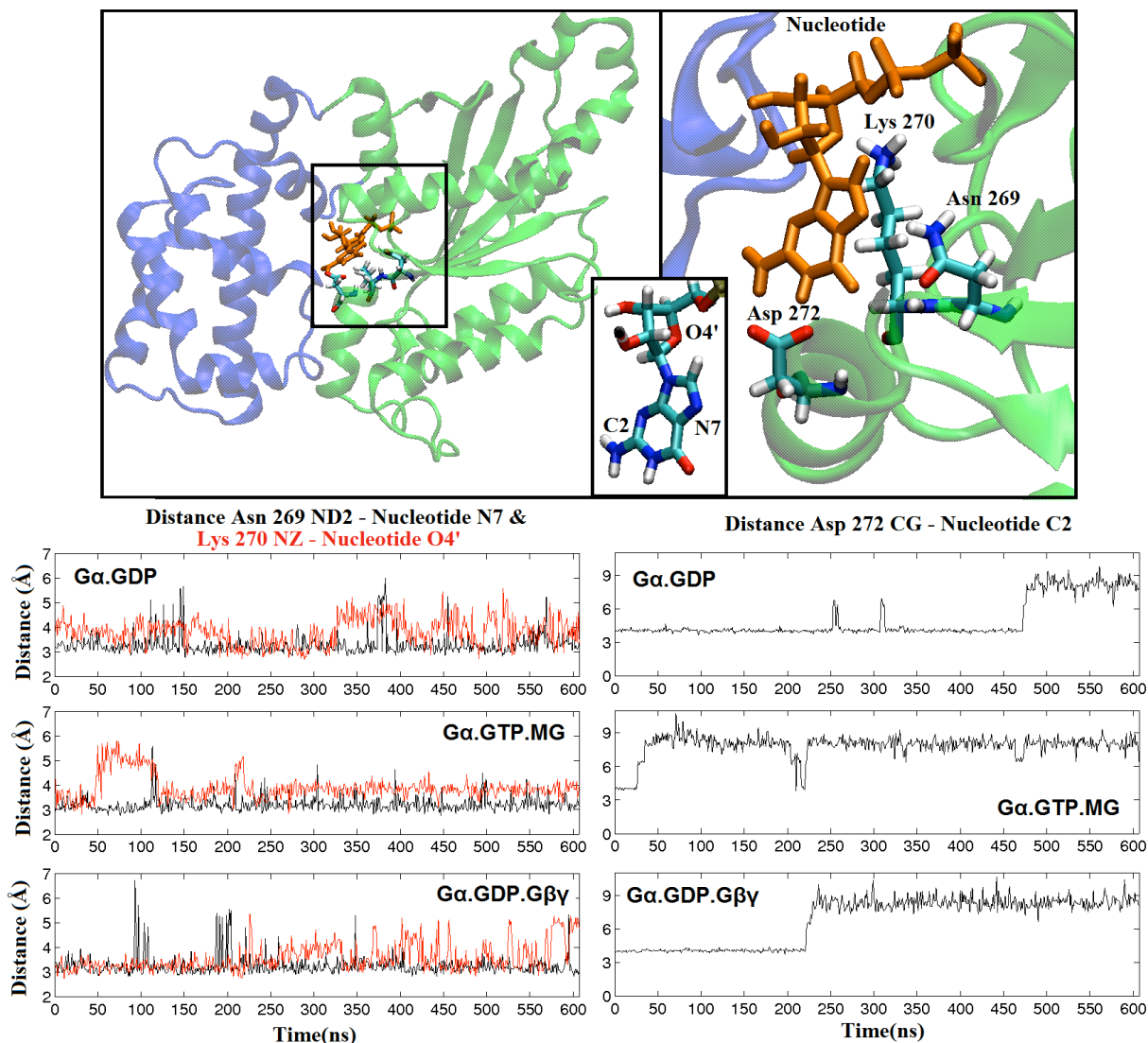


Figure 5.6. Interactions in binding group five. Top left, location of binding group 5 in the α -subunit with the GTPase domain in green and the α helical domain in blue. Top right, Asp 272, Asn 269 and Lys 2700 interacting with the nucleotide. Top inset, labeling of atoms in the nucleotide that interact with the side chains in this group. Bottom left, distances from Asn 269 ND2 to the Nucleotide atom N7; and from Lys 270 NZ to the Nucleotide atom O4'. Bottom right, distances from Asp 272 CG to the Nucleotide atom C2, and from Arg 176 HH22 to Nucleotide N2.

The last important residue in this group is Asp 272. It forms a double hydrogen bond with the guanine group of the nucleotide in a way that it appears to lock the nucleotide in place. In $G_{\alpha.GDP}$, the hydrogen bonds break and re-associate a few times before dissociating for good after 470 nanoseconds into the simulation. The opposite occurs in $G_{\alpha.GTP.MG}$, the dynamics of the hydrogen bond changes as it starts associated and then dissociates 25 nanoseconds after the

simulation begins. The double hydrogen bond is also formed at 209 and 215 to 220 nanoseconds; This double bond is less conserved in the active monomer because of the additional phosphate group in GTP that pushes it away towards the solvent. Finally $G_{\alpha.GDP}.G_{\beta\gamma}$ forms the double hydrogen bond with the nucleotide for the first 220 nanoseconds before dissociating for the remaining of the simulation. Asp 272 is an example of a residue that departs from the position observed in the crystal structure.

Table 5.6. PB nucleotide mutation binding energies for binding group five.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Asn 269	-3.92	-2.60	-4.54	1.84	1.32	-8.46
Lys 270	-36.38	-25.27	-31.80	6.53	-11.11	4.58
Asp 272	-5.09	-0.42	-1.06	0.64	-4.67	4.03
Crb 269	-2.12	-1.01	-1.65	0.64	-1.11	0.47
Crb 270	-1.69	-0.80*	0.00	-0.80*	-0.89	1.69*
Amn 270	-1.68	-0.50	0.01	-0.51*	-1.18	1.69*

^a(A) $G_{\alpha.GDP}$, (B) $G_{\alpha.GTP.MG}$ and (C) $G_{\alpha.GDP}.G_{\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP}.G_{\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP}.G_{\beta\gamma}) - \Delta G(G_{\alpha.GDP})$. * Energies calculated with GB. Crb stands for carbonyl group, and Amn for amino group. All units in kcal/mol.

5.4.1.2. Binding between G_{α} and $G_{\beta\gamma}$

This binding interface is large compared to the nucleotide's, and it has an average of $495.93 \pm 53.28 \text{ \AA}^2$. It has fewer components than nucleotide binding for components within a 1 kcal/mol cutoff. There are 28 components interacting at the binding interface with a cutoff of 1 kcal/mol. Of those components, 15 belong to $\beta\gamma$, and 13 to α . Of these components, only four belong to the backbone: carbonyl 16 and carbonyl 213 in the α -subunit, and carbonyl 117 and carbonyl 118 in the $\beta\gamma$ subunit. Figure 5.7 shows a binding profile of every side chain in the heterotrimer protein-protein interface.

There are three possible reasons for having fewer components despite a larger binding surface. One is that the interface is not as tight as the one by the nucleotide wrapped inside a α subunit. Another one is that the interactions are made between components and not between a component and a nucleotide that has larger partial charges. Lastly, the barrel shape of $\beta\gamma$ has a hollow center and interactions between the two subunits are only possible around the side of the

barrel shape structure that is closer to α , and along the side of the $\beta\gamma$ barrel that is in contact with the alpha helical region of the N-terminus of α .

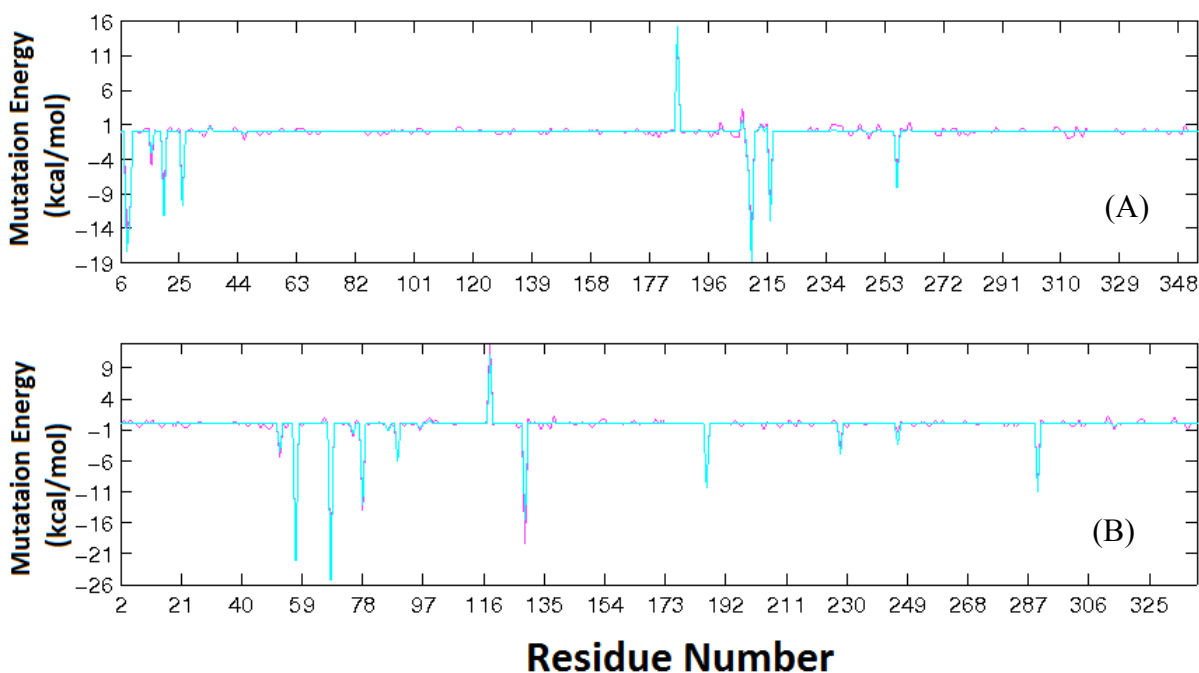


Figure 5.7. Contributions to protein–protein binding. PB (cyan) and GB (pink) heterotrimer binding for the α and $\beta\gamma$ binding interface. (A) Side chain components corresponding to the α sub-unit (B) Side chain components corresponding to the β sub-unit.

The main side-chain components in the protein–protein heterotrimer interface can be grouped according to their proximity along the $\alpha\beta$ binding interface; this facilitates the independent analysis of each in the remaining of this sections.

5.4.1.2.1. Protein–protein binding group one

This group is near the α -subunit N-terminus. In the crystal structure, its components do not show any interaction, but during the simulation, Asp 8 (-17.4 kcal/mol) and Asp 9 (-12.9 kcal/mol) in the α -subunit interact favorably with Arg 68 (-25.4 kcal/mol), Arg 129 (-15.7 kcal/mol) and Thr 86 (-1.3 kcal/mol) in the β subunit. Figure 5.8 shows the location of the group in the heterotrimer as well as the two predominant states.

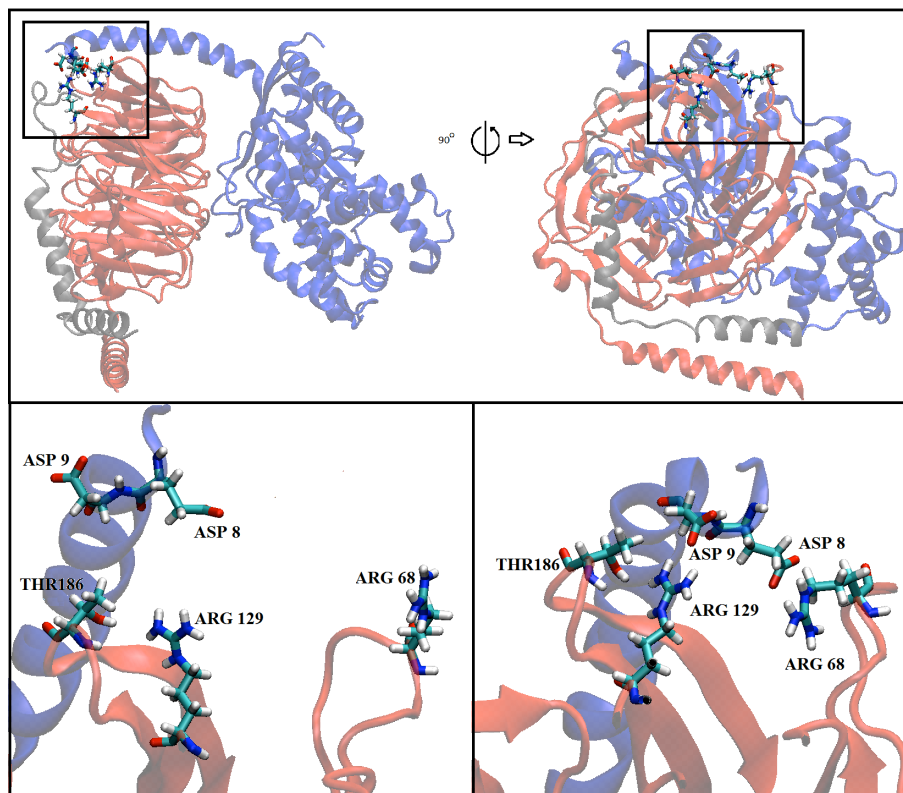


Figure 5.8. Interactions protein–protein group one. Location and main conformations of protein–protein binding group one. Above, location of the group in the $\alpha\beta$ interface. Below left, initial state of the group at the beginning of the simulation. Below right, sample frame at 181 nanoseconds when the residues in this group interact with each other.

The side chains in this group start interacting with each other after about 70 nanoseconds into the simulation. These interactions remain stable for the rest of the simulation, with only brief separations due to water molecules in their surroundings that might interfere. For the most part, Asp 9 in α and Thr 186 and Arg 129 in β interact with each other. Also, Asp 8 in α and Arg 68 in β interact mostly with each other. Asp 8 mediates between Arg 68 and Arg 129, in β , for brief periods of time during the simulation. While the interactions between the residues in this group are generally conserved, they are in one end of the protein which allows them to move broadly, and their biological interpretation is difficult to make knowing that this end of the protein is restrained by post-translational modifications that insert the cell membrane.

5.4.1.2.2. Protein–protein binding group two

The residues in this group are also located at the interface formed by the N-terminus α helical region of α that is extended along the β -subunit. Figure 5.13, shows this group formed by

Asp 26 (-10.8 kcal/mol), Asp 20 (-12.3 kcal/mol) and Ser 16 (-3.2 kcal/mol) in the α -subunit, and Lys 78 (-13.1 kcal/mol), Arg 52 (-4.5 kcal/mol), Lys 89 (-6.2 kcal/mol) in the β -subunit. Asp 26 and Arg 52 would only interact at 1 nanosecond and between 120 and 232 nanoseconds out of the whole simulation. Lys 89 does not interact so cleanly with Asp 20 and Ser 16, they dissociate intermittently, but they tend to be associated around the 120 to 232 nanoseconds interval of Asp 26 and Arg 53. Asp 26 in α and Lys 78 in β have a more conserved interaction with no long dissociations, only brief intermittent ones, also, as seen in Figure 5.9.

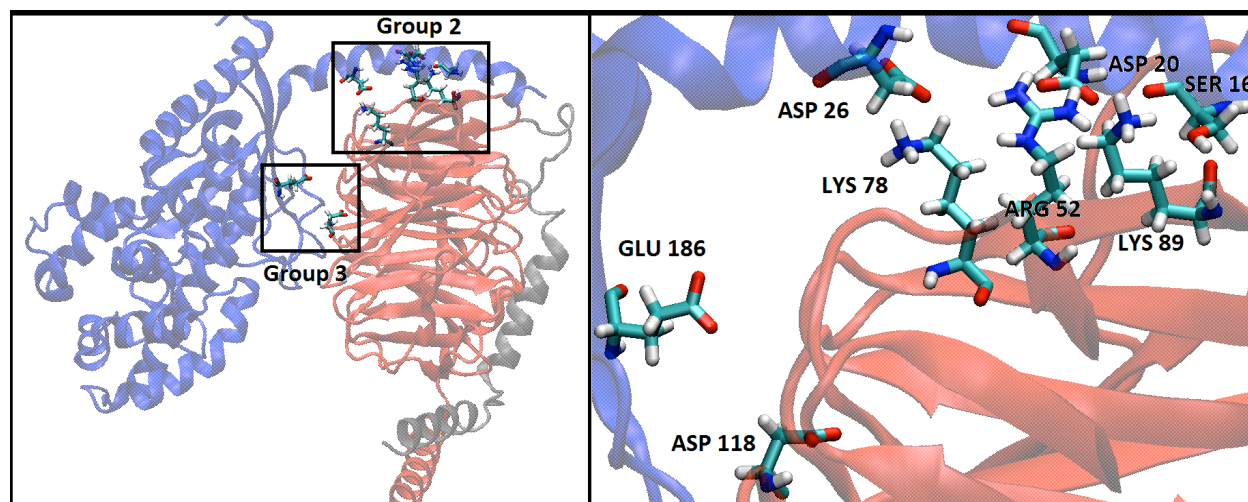


Figure 5.9. Interactions protein–protein group two. Left, location of protein–protein binding groups 2 and 3 in the $\alpha\beta$ interface. Right, closer view of the groups showing interactions among its residues.

5.4.1.2.3. Protein–protein binding group three

Figure 5.9 also shows the interactions between Glu 186 (15.2kcal.mol) in α and Asp 118 (11.2kcal/mol) in β ; this interaction is unfavorable—they never get closer than 3 Å—and suggests that some unfavorable interactions are necessary to facilitate the dissociation of a large binding interface.

5.4.1.2.4. Protein–protein binding group four

Figure 5.10 shows the group four location along the $\alpha\beta$ binding interface on the right, in the top black block. The block in the top left shows a closer view of residues Glu 216 (-13.0 kcal/mol) in α , and Lys 57 (-22.2 kcal/mol) and Gln 75 (-1.6 kcal/mol) in β . Gln 75 tends to interact with the backbone and Lys 57 with the side chain of Glu 216. These two interactions

move a lot, making it difficult to describe in terms of dissociation and association of hydrogen bonds and side-chain flips in Gln 75.

5.4.1.2.5. Protein–protein binding group five

Figure 5.10 shows the group five location along the $\alpha\beta$ binding interface on the right, in the middle black block. The block in the middle left shows a closer view of residues Lys 257 (-8.2 kcal/mol) in α and Asp 290 (-11.1 kcal/mol) in β . While the interaction of these two residues is also very noisy, with salt bridges formed and broken constantly, the two side chains start far apart for the most part in the first 200 nanoseconds, and seem to form more frequent and closer interactions for the remaining of the simulation. A similar pattern is observed for the interactions between Lys 257 and Carbonyl 290.

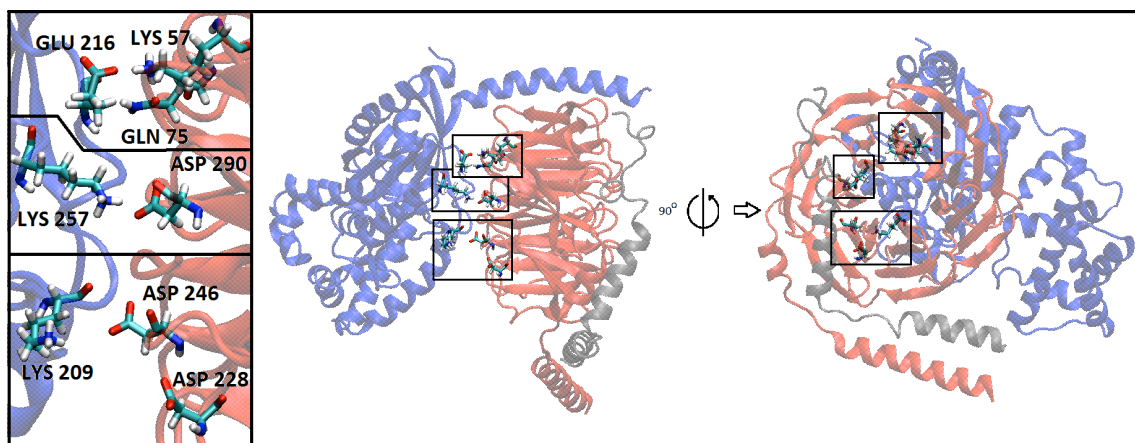


Figure 5.10. Interactions in protein–protein group 4, 5 and 6. Left, zoom in on the groups showing interactions among its residues. Left, from top to bottom: protein–protein binding group 4; 5 and 6. Right, location of protein–protein groups 4, 5 and 6 in the $\alpha\beta$ interface from two different perspectives.

5.4.1.2.6. Protein–protein binding group six

Figure 5.10 shows the group six location along the $\alpha\beta$ binding interface on the right, in the bottom black block. The block in the bottom left shows a closer view of residues Lys 209 (95.8 kcal/mol) in α , and Asp 246 (-3.3 kcal/mol) and Asp 228 (-5.1 kcal/mol) in β . Lys 209 interacts more strongly with both Asp 246 and 228 after about 150 to 175 nanoseconds. Asp 246 forms a stronger interaction while Asp 228 dissociates between 400 and 475 nanoseconds and for the last 80 nanoseconds of the simulation. These interactions are too noisy to characterize in any more detail.

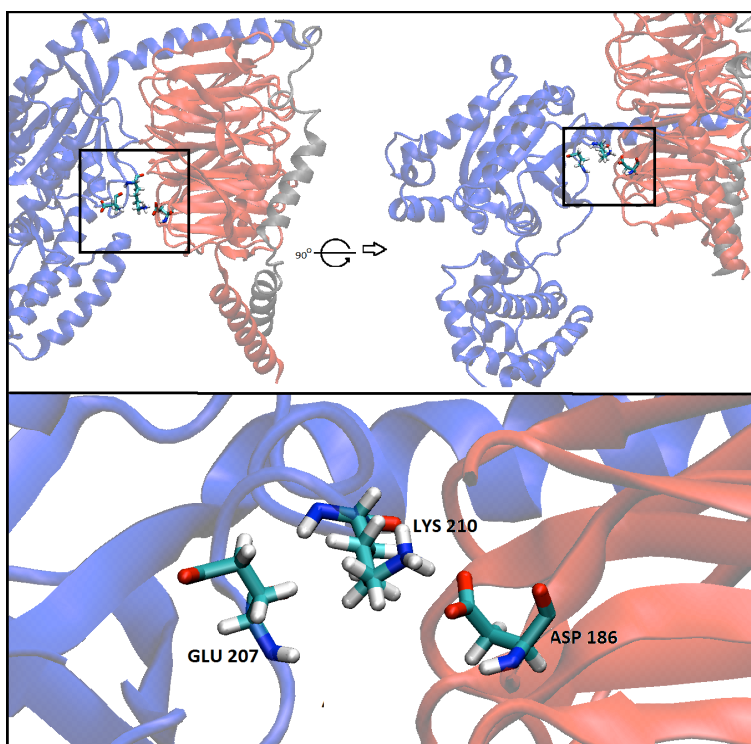


Figure 5.11. Interactions in protein–protein group 7. Top, location of the protein–protein binding group 7 in the $\alpha\beta$ interface. Bottom, zoom in on the group showing interactions among its residues.

5.4.1.2.7. Protein–protein binding group seven

Figure 5.11 shows groups seven location along the $\alpha\beta$ binding interface on the right, in the top black block. In the bottom, Glu 207 (1.7 kcal/mol) and Lys 210 (-18.8 kcal/mol) in α , and Asp 186 (-10.4 kcal/mol) in β . Lys 210 and Asp 186 do not interact until after 195 nanoseconds into the simulation. Glu 207 does not interact with Lys 210 until 217 nanoseconds into the simulation; this interaction is very unstable and breaks and forms intermittently, without any clear patterns throughout the remaining of the simulation. In this mode of interaction, Lys 210 is an intermediary between two negatively charged residues, one in its own chain and the other one in a different chain across the binding interface.

With group seven ends the analysis of the interactions in the $\alpha\beta$ interface. The following analysis is done on residues that are important in folding, and that have not already been analyzed structurally in binding; as some residues have both binding and folding interactions.

5.4.2 Important components for folding

Important components for folding are those that have a significant electrostatic interaction with other components besides the nucleotide. There are still some residues with low binding energies that interact with residues that interact with the nucleotide; in order to identify these residues from the simulation of proteins with many components, the calculation of folding energies is a useful tool. Folding calculations correspond to the energy of transferring every residue side chain in the protein from the solvent to the folded state in the protein. In this way, important interactions between side chains emerge in the rankings according to energy. For example, a residue that forms a salt bridge inside the protein would have favorable transfer energy from the solvent to its environment in the protein. In this way, networks of salt bridges emerge that interconnect important regions of the G-protein in ways that give insights into their potential relevance in signal transduction through the G-protein. Folding gives an idea of how the exposure to solvent changes through the simulation. For example, a side chain that is kept isolated from the solvent through the simulation would give a large energy of folding; a side chain that remains completely solvated through the simulations would give zero folding energy.

As in Figure 5.1 for binding, folding mutation energies of side chains are shown in Figure 5.12 on the left. This figure can be seen as an electrostatic profile that shows residues energies of

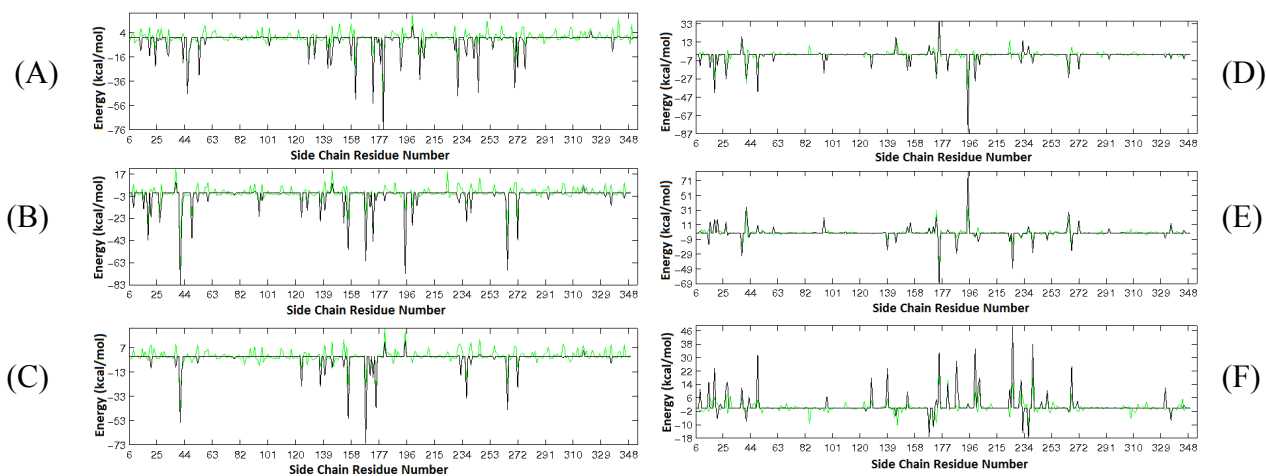


Figure 5.12. Contributions to folding free energy. PB(black) and GB(green) side chain mutation energies for folding of G_{α} to the nucleotide for selected components by GB. Left, (A) $G_{\alpha.GDP}$. (B) $G_{\alpha.GTP.MG}$ (C) $G_{\alpha.GDP.G\beta\gamma}$. Right, (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP.G\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP.G\beta\gamma}) - \Delta G(G_{\alpha.GDP})$.

folding for each amino acid side chain. Some of the components significant in folding for the three systems have conserved interactions that do not change from one state of the G-protein to another. In order to determine residues with relevant changes, the plots in Figure 5.12 A, B and C, on the left, were subtracted from one another to give side chain differences in folding mutation energies from one state of the heterotrimeric G-protein to the other; these are shown in Figure 5.12 on the right.

5.4.2.1. Folding group one

This group shows significant folding energies, and it is located near the other loop region that connects the GTPase and α helical domains. It is formed by Lys 54, Glu 58 and Glu 65. The three residues interact forming a double salt bridge. The salt bridge between Lys 54 and Glu 65 is very loose in the heterotrimer in comparison to the two monomers; this salt bridge appears to hold the GTPase and α helical domain in the monomers but is more unstable in the heterotrimer, distance plots in Figure 5.13. It has been suggested that conformational motions between this two domains is important in nucleotide exchange, and that this nucleotide exchange is initiated when the G-protein is in the heterotrimeric state.(21) The breaking of this salt bridge would contribute to this mechanism.

Table 5.7 shows that Lys 54 has a favorable folding energy for the monomers, and not very significant for the heterotrimer. The table also shows that going from the heterotrimer to the monomer with a GTP is very favorable, as it would be expected for such transition since it would favor the formation of the salt bridge. Going from the inactive monomer to the heterotrimer is unfavorable as it would cause the salt bridge to be broken.

Table 5.7. PB mutation folding energies for group 1, and change in folding energies from one system to the other.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Lys 54	-31.33	-41.29	-5.34*	-41.29	9.96	25.99*
Glu 58	-6.12	-8.64	-5.98	-2.66	2.52	0.14
Glu 65	0.64*	-8.53	6.14*	-14.67*	9.17*	5.5*

^a $G_{\alpha.GDP}$, (B) $G_{\alpha.GTP.MG}$ and (C) $G_{\alpha.GDP.G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP.G\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP.G\beta\gamma}) - \Delta G(G_{\alpha.GDP})$. * Energies calculated with GB. Crb stands for carbonyl group and Amn for amino group. All units in kcal/mol.

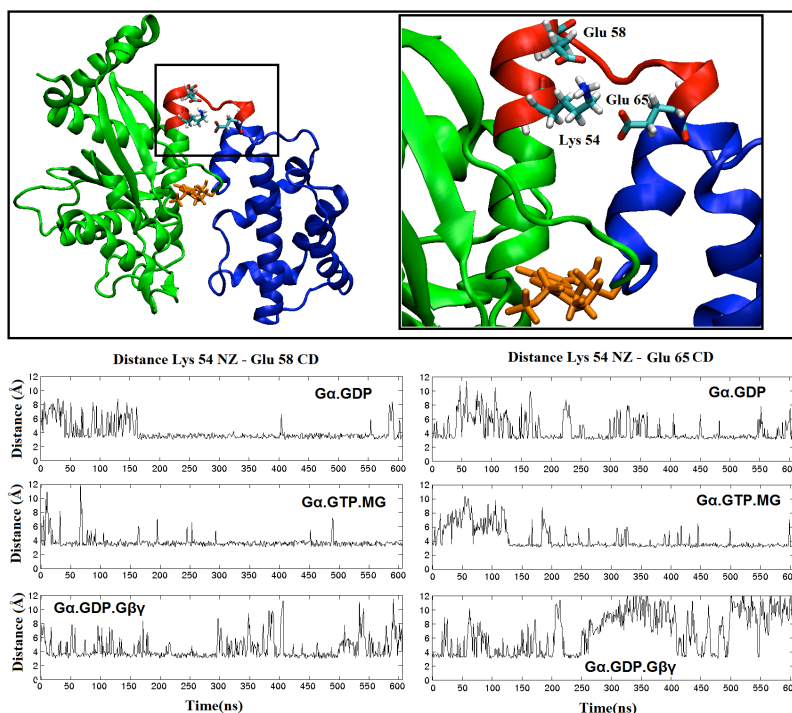


Figure 5.13. Interactions in folding group one. Schematics and distances between heavy atoms in folding group 1. Top, this loop is located near the region of the protein that connects the GTPase domain and the α helical region. In green, the GTPase domain; in red, the loop region between Lys 54 and Glu 65; in blue, the α helical domain. Bottom, distances between heavy atoms in forming the salt bridge, left, distance between Lys 54 NZ and Glu 58 CD, right, distance between Lys 54 NZ and Glu 65 CD.

5.4.2.2. Folding group two

This group is positioned away from the GTPase domain and the nucleotide. From its location, it does not appear to play a direct role in nucleotide exchange or trimer dissociation, Figure 5.14. It is believed that for nucleotide exchange to occur, the α helical domain opens relative to the GTPase domain; if this is the case, then this group of residues may interact with other proteins to facilitate this process.

This group of residues may change the way they interact among themselves when the G-protein is in the trimeric, inactive state. In Table 5.8, Asp 102 and Asp 133 do not show any folding interaction when the $\beta\gamma$ subunit is bound to the GTPase domain, but they show favorable folding energies between -15.96 and -22.50 kcal/mol for the two monomers. In other words, in the trimeric state these two residues interact with the solvent for most of the simulation.

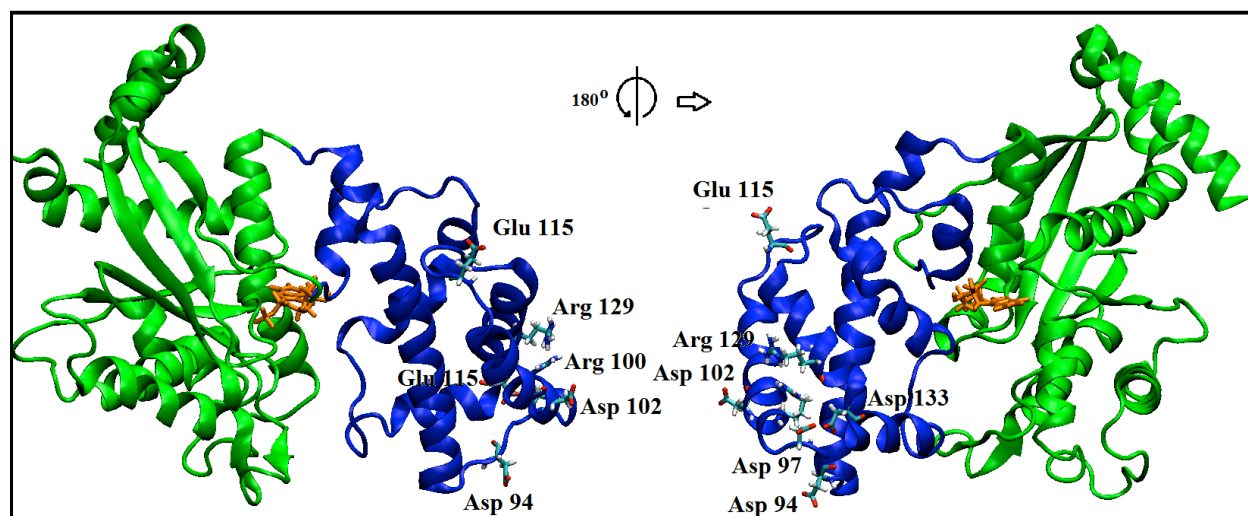


Figure 5.14. Interactions in folding group two. Schematics of folding group 2. This group is located in the α helical region that is far from the GTPase domain and the nucleotide. Because of its location, it does not appear to play a role in heterotrimeric G-protein function, but it could interact with signaling downstream effectors or other G-protein interacting proteins that help in its transitions through the cycle.

Table 5.8. PB mutation folding energies, and change in folding energies for one system to the other for folding group two.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Asp 94	0.32	-0.44	6.78*	-7.22*	0.44	6.46*
Asp 97	0.97	1.50	8.05*	-6.55*	-0.53	-0.97*
Arg 100	-0.37*	-21.48	-8.13*	29.61*	21.48	-7.76*
Asp 102	-6.78	-7.04	7.64*	-14.68*	0.26	14.42*
Glu 115	-0.41	-1.12	6.68*	-7.80*	1.12	7.09*
Arg 129	-22.50	-22.39	-24.71	2.32	-0.11	-2.21
Asp 133	-17.83	-15.96	-7.48*	-8.48*	-1.87	10.35*

^a $G_{\alpha,GDP}$, (B) $G_{\alpha,GTP,MG}$ and (C) $G_{\alpha,GDP,G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha,GTP,MG}) - \Delta G(G_{\alpha,GDP,G\beta\gamma})$. (E) $\Delta G(G_{\alpha,GDP}) - \Delta G(G_{\alpha,GTP,MG})$. (F) $\Delta G(G_{\alpha,GDP,G\beta\gamma}) - \Delta G(G_{\alpha,GDP})$. * Energies calculated with GB. Crb stands for carbonyl group and Amn for amino group. All units in kcal/mol.

5.4.2.3. Folding group three

This group shows the most important non-covalent interactions between the GTPase domain and α helical domain; it is away from the two loop regions that connect this subdomains. If the GTPase and the α helical domain separate to exchange nucleotides, the salt bridge formed between Arg 144, Asp 229 and Asp 231 would break. The distances between the constituents of this salt bridge can be seen in the bottom left of Figure 5.15. In this figure, the

active monomer forms the salt bridge for only a few nanoseconds. The inactive monomer shows that the salt bridge forms sporadically until about 250 nanoseconds into the simulation, when the interactions become well conserved. In the heterotrimer the salt bridge is formed between about 50 and 90 nanoseconds; at any other time, the salt bridge is loose. It is important to note the differences in distance between Asp 231 and Arg 144 in $G_{\alpha.GDP}.G_{\beta\gamma}$ and $G_{\alpha.GDP}$ when the salt bridge is formed; it is a lot closer for $G_{\alpha.GDP}$.

The interactions of Lys 277 with Asp 229 and 231 show no correlation with the formation of the salt bridge between of the two aspartates and Arg 144 for the monomers. For the heterotrimer, the breaking of the salt bridge of the aspartates with Lys 277 appears to coincide with the formation of the salt bridge with Arg 144.

Table 5.9. PB mutation folding energies, and change in folding energies for one system to the other for folding group three.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Arg 142	-25.68	-25.88	-24.82	-1.07	0.21	0.86
Arg 144	-23.75	-0.34*	-1.87*	1.53*	-23.75	-21.81*
Glu 145	-16.56	-16.45	-16.05	-0.41	-0.11	0.52
Asp 229	-11.40	1.31	4.00*	1.31	-12.71	15.40*
Asp 231	-48.73	0.36*	-4.02*	4.38*	-49.09*	44.71*
Asp 277	-26.02	-42.87	-26.05	-16.82	16.85	-0.03

^a $G_{\alpha.GDP}$, (B) $G_{\alpha.GTP.MG}$ and (C) $G_{\alpha.GDP}.G_{\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP}.G_{\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP}.G_{\beta\gamma}) - \Delta G(G_{\alpha.GDP})$. * Energies calculated with GB. Crb stands for carbonyl group and Amn for amino group. All units in kcal/mol.

The energies of interaction for this group are shown in Table 5.9. Arg 144 only has a significant folding energy in the inactive monomer, where the salt bridge is form after 250 nanoseconds. It is notable that Arg 144, together with Asp 229, forms the salt bridge for about 140 nanoseconds in the heterotrimer, yet the folding energy average for the whole simulation is close to zero.

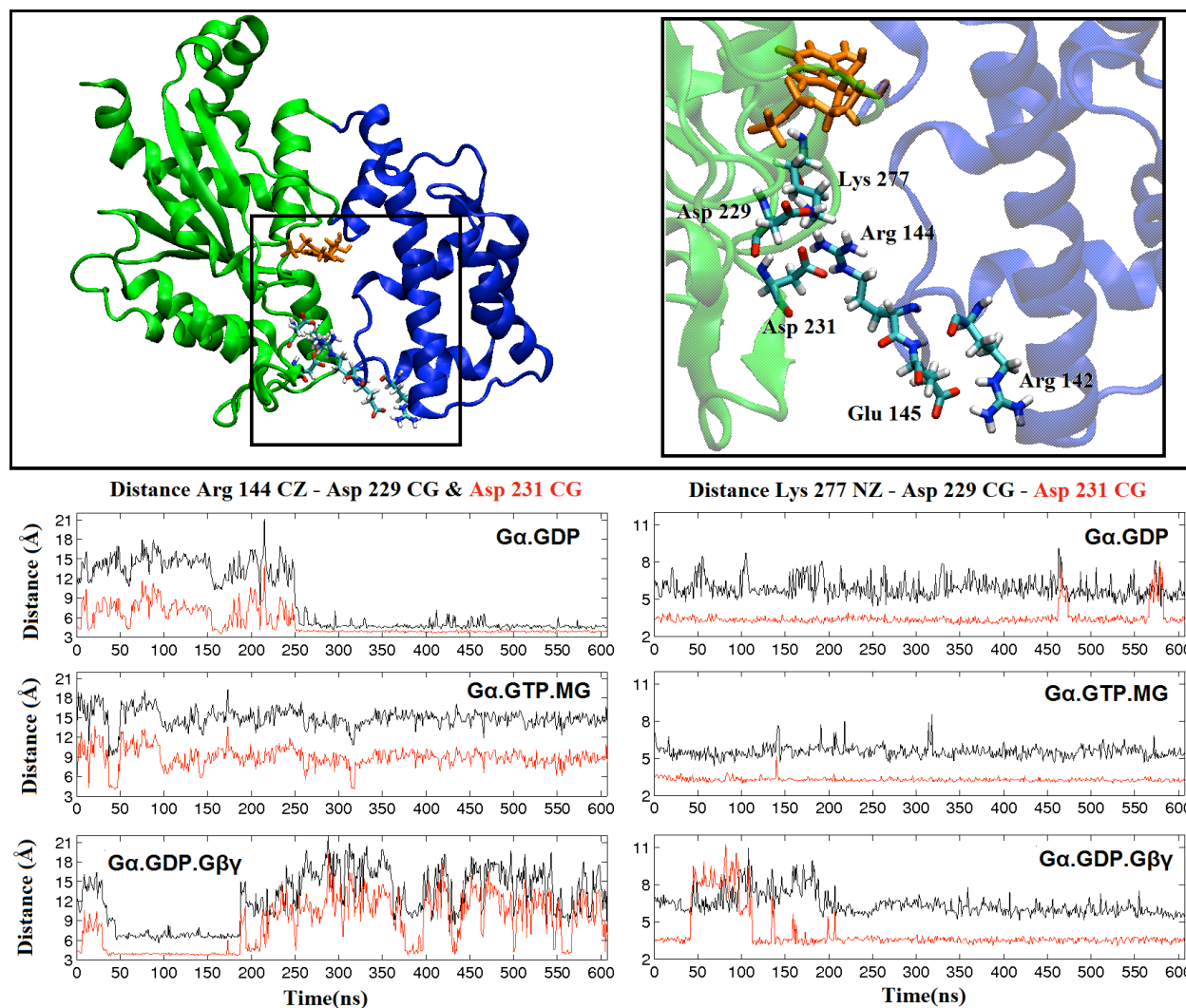


Figure 5.15. Interactions in folding group three. Schematics and distances between heavy atoms in folding group 3. Top, this group of residues is located opposite, relative to the nucleotide, from switch region one and the loop region formed by folding group 1. In green, the GTPase domain; in blue, the α helical domain. Bottom, distances between heavy atoms forming the salt bridge, left, distances between Arg 144 CZ, Asp 229 CG and Asp 231 CG. Right, distances between Lys 277 NZ and Asp 229 CG and Asp 231 CG.

There is another salt bridge formed within the α subunit only between Arg 142 and Glu 145. This salt bridge is so conserved in the three systems that the average distance between the CZ atom in Arg 142 and the CD in Glu 145 is the exactly the same, 3.92 Å. The energies of folding of these two components are also very similar in the three systems suggesting that this salt bridge plays no significant role in the heterotrimeric G-protein in its different states despite of its location close to the interface between the GTPase domain and the α helical domain.

5.4.2.4. Folding group four

The interactions in this group modulate the motions of switch regions two and three. Switch region two is in contact with the $\beta\gamma$ subunit, and it has more mobility in $G\alpha.GDP$ than $G\alpha.GTP.MG$. The residues that form the salt bridges in this group do not interact with the $\beta\gamma$ via a hydrogen bond or salt bridge. They seem to modulate the degree this group moves in the three different states of the heterotrimeric G-protein.

One of the salt bridges is formed between Arg 205 and Asp 236, Glu 237 and Asp 238. Figure 5.16 shows that the distance between three carboxylic group in the negatively charge residues and the arginine varies simultaneously despite some of the components in switch region three being farther. It is possible to see jumps in Asp 238 and 236 at similar times in the simulation. This salt bridge is more stable in the inactive monomer than in the active monomer, as seen in the bottom left plots of Figure 5.16. The structural backbone analysis of switch region moves more in the inactive monomer. Also, the crystal structure of the inactive monomer does not show diffraction for switch region two, yet the salt bridge between switch region two and three is more conserved in the inactive monomer. An explanation for this is that the increase mobility for switch region one in the monomer is due to internal motions of the whole GTPase domain that are transmitted to it through switch region three. If differences in mobility between the two regions were due solely to the solvent they would move about the same. Switch region three does not show a significant difference in mobility between the two monomers, suggesting that its mobility is coupled to the GTPase domain to a greater extent than to switch region two. The dependence of mobility of switch region two to switch region three, and in turn to the GTPase, domain, is an interesting observation that links this salt bridge to a nucleotide that is harbored away from it. This suggest some interconnectivity that supports changes in roles for the GTPase domain when different nucleotides are bond to it, and this changes can be attributed to changes in *behavior* between the three states of the heterotrimeric G-protein.

The other salt bridge of importance in this group is between two arginines and a glutamate. This salt bridge interacts differently for the three states of the G-protein. While this salt bridge is close to the salt bridge discussed in the previous paragraph, and has interactions with it, it interacts differently in the three systems, as seen in Figure 5.17. The tighter interaction

between Arg 205, which is part of a helical region of the GTPase domain, in the inactive monomer supports the notion that the increased mobility of switch region one in the inactive monomer is due to its stronger connection with the GTPase domain through salt bridges.

There no salt bridge formed between Glu 238 and Arg 242 in any of the systems. The distance between the carbon atoms at the center of their side chains do tell something about the differences in these regions between the different states of the heterotrimeric G-protein. The average distance between the CD atom in glutamate and the CZ atom in arginine is similar in the two monomers, 15.09 ± 1.17 Å for the inactive monomer and 15.08 ± 1.51 Å for the active monomer, and closer in the heterotrimer, 9.96 ± 0.68 Å.

Table 5.10. PB mutation folding energies, and change in folding energies for one system to the other for folding group 4.^a

	Binding Energy (ΔG)			Difference in binding ($\Delta\Delta G$)		
	A	B	C	D	E	F
Arg 205	-35.39	-29.69	-5.34*	-24.35*	-5.70	30.05*
Arg 208	-17.91	-11.77	0.16*	-11.93*	-6.17	18.07*
Glu 236	-8.98	-5.49	7.47*	-12.96*	-3.49	16.45
Asp 237	-15.68	-9.49	1.14	-10.59	-6.23	16.82
Glu 238	-0.84	-0.58	-15.81	15.23	-0.26	-14.97
Arg 242	-17.37	-26.40	-34.97	8.57	9.03	-17.60
Glu 245	-45.74	-18.59	-7.82	-10.76	-27.16	37.92

^a $G_{\alpha.GDP}$, (B) $G_{\alpha.GTP.MG}$ and (C) $G_{\alpha.GDP.G\beta\gamma}$. $\Delta\Delta G$ mutation energies between systems: (D) $\Delta G(G_{\alpha.GTP.MG}) - \Delta G(G_{\alpha.GDP.G\beta\gamma})$. (E) $\Delta G(G_{\alpha.GDP}) - \Delta G(G_{\alpha.GTP.MG})$. (F) $\Delta G(G_{\alpha.GDP.G\beta\gamma}) - \Delta G(G_{\alpha.GDP})$. * Energies calculated with GB. Crb stands for carbonyl group and Amn for amino group. All units in kcal/mol.

The folding energies for residues in this group are in Table 5.10. Arg 205, Arg 208, Glu 236 and Asp 237 have favorable folding energies for the two monomers only and very low for the heterotrimer. Glu 238 has an opposite interaction profile, it is more favorable in the heterotrimer and not very significant for the monomers. Arg 242 and Arg 245 have favorable folding energies in the three systems although their variations are in the tens of kilo calories per

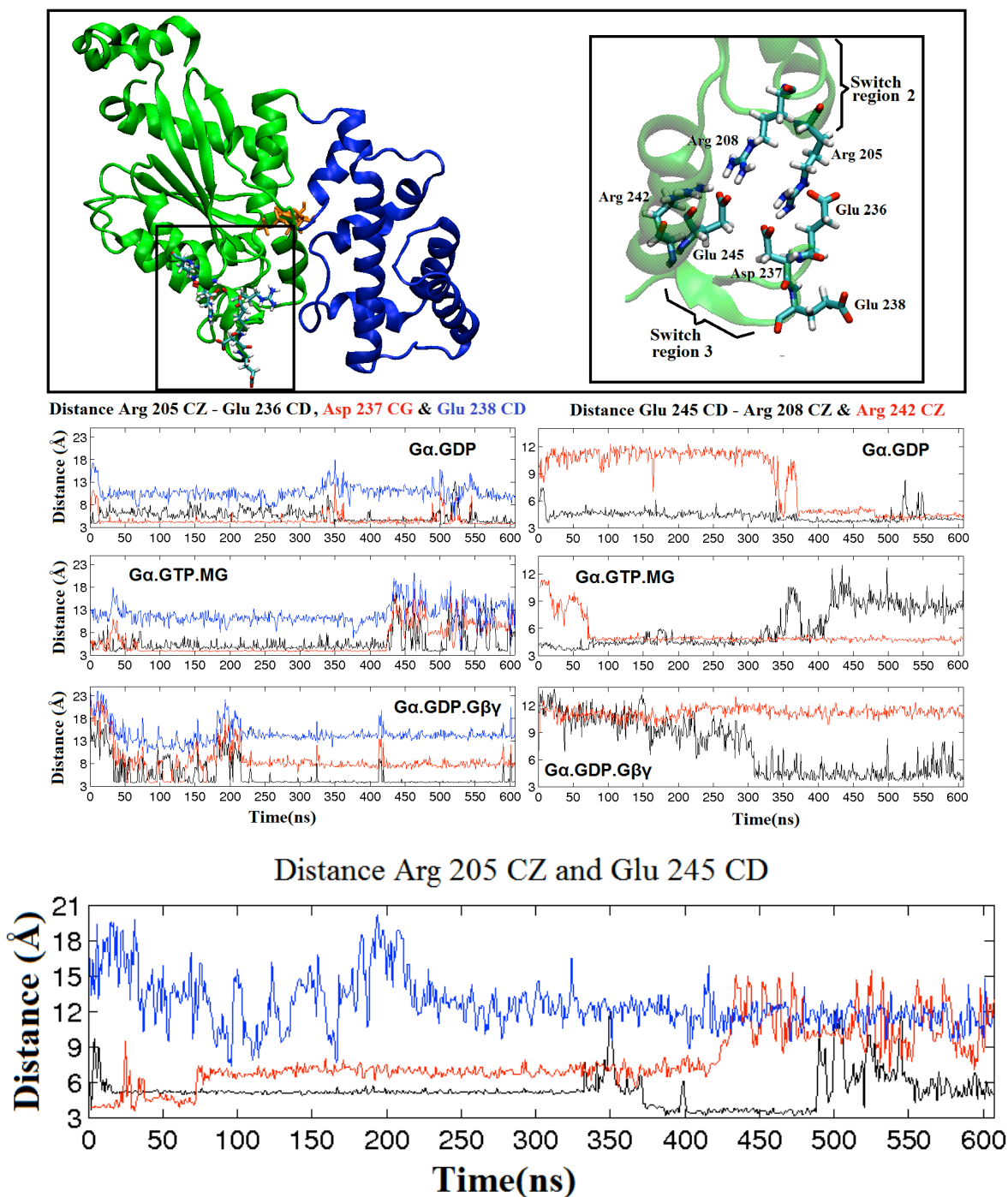


Figure 5.16. Interactions in folding group 4. Schematics and distances between heavy atoms in group 4. The box on the left is zoomed in in a different angle on the right to better show structural features. Top, this group of residues interconnect switch region two and switch region three. In green, the GTPase domain; in blue, the α helical domain. Middle, distances between heavy atoms forming the salt bridge, left, distances between Arg 205 CZ, Glu 236 CD, Asp 237 CG and Glu 238 CD. Right, distances between Glu 245 CD and Arg 208 CZ and Arg 242 CZ. Bottom, distance between Arg 205 CZ and Glu 245 CD of folding group 4. Black, the inactive monomer $G_{\alpha.GDP}$. Red, the active monomer $G_{\alpha.GTP.MG}$. Blue, the heterotrimer $G_{\alpha.GDP.G\beta\gamma}$.

mol. There are very high transition energies of folding between states for this group region, as seen in columns D, E and F, but no trend can be derived from these transitions because of the intricate electrostatic interactions that couples them; however, the significant changes in this region, and the discrete explanation of interactions, support that there are relevant allosteric interactions due to nucleotide exchange alone.

5.5 Conclusion

The structural characterization of components relevant to binding and folding was studied in this chapter. From the ranking of mutation energies for individual components, it is easier to identify important components that may be relevant for function in the heterotrimeric G-protein. Going into details of side chain conformations show clear differences in the three states that can be attributed to differences in nucleotide or the presence of the $\beta\gamma$ subunit. These differences, such as in Glu 43 and Asp 272, indicate structural determinants that can be related to function.

Important components in binding and folding were subdivided by proximity into interacting groups. For the most part, the interactions analyzed were between residues in the group. These groups sometimes spawn several regions that are not necessarily contiguous in sequence. For example, folding group four is formed by residues in both switch region two and three. In this way, component analysis of folding and binding free energies provides alternative way to divide the α sub-unit of the heterotrimeric G-protein that is different than the usual division based on contiguous sequential blocks.

This analysis gives a more detailed look into function of the heterotrimeric G-protein than can be given by analyzing the crystal structures alone. The energetic and structural decomposition can be done on the crystal structure alone if the theoretical molecular dynamics are not used because of their inherent uncertainty. If dynamic information from simulations is considered, additional information could be inferred to help understand G-protein function.

CHAPTER 6

Conclusions and future directions

6.1 Conclusions

The study of the heterotrimeric G-protein from theoretical models based on molecular dynamics simulations gives deeper insight into the mechanisms of signal transductions at the molecular and atomic level. This insight does not emerge from direct observation of the molecular dynamic trajectories but from the systematic analysis of its structures and energetics. The emergence of determinants of functions reveal that signal transduction through oligomers with multiple sub-domains is not a disordered mechanism; instead, it is carefully concerted through interactions between loop regions, other secondary structural features, multiple helix domains and nucleotides. The exact way these concerted motions occur is not fully understood; these molecular dynamic simulations and their analysis give insights into concerted mechanisms of heterotrimeric G-protein function.

The analysis shown here includes regions that are considered of importance by the crystallographic papers from which the initial trajectories are obtained for the simulation. The relaxation of switch regions II and III from the active to inactive states is mentioned in the crystallographic papers to be important in the interaction between the α and β subunit as well as nucleotide exchange.(16-18) In particular, Coleman et. al. (17) suggest the importance of Glu 204 and Arg 178; both residues identified by energetic component analysis to be important in Chapters 3 and structurally characterized in Chapter 5. Another section of importance studied is the helical bundle that is attached to the GTPase domain which is believed to move in such way as to contribute in nucleotide exchange.(21) The ranges of motions described in Chapter 4 for this helical region shows clear differences in motions for the three different states of the G-protein α -subunit.

As a conclusion of this work, the energetic information can be summarized with the aid of primary structure maps. Figure 6.1 shows the primary structure map for the three systems studied in this thesis. One application of the primary structure map in Figure 6.1 is to identify

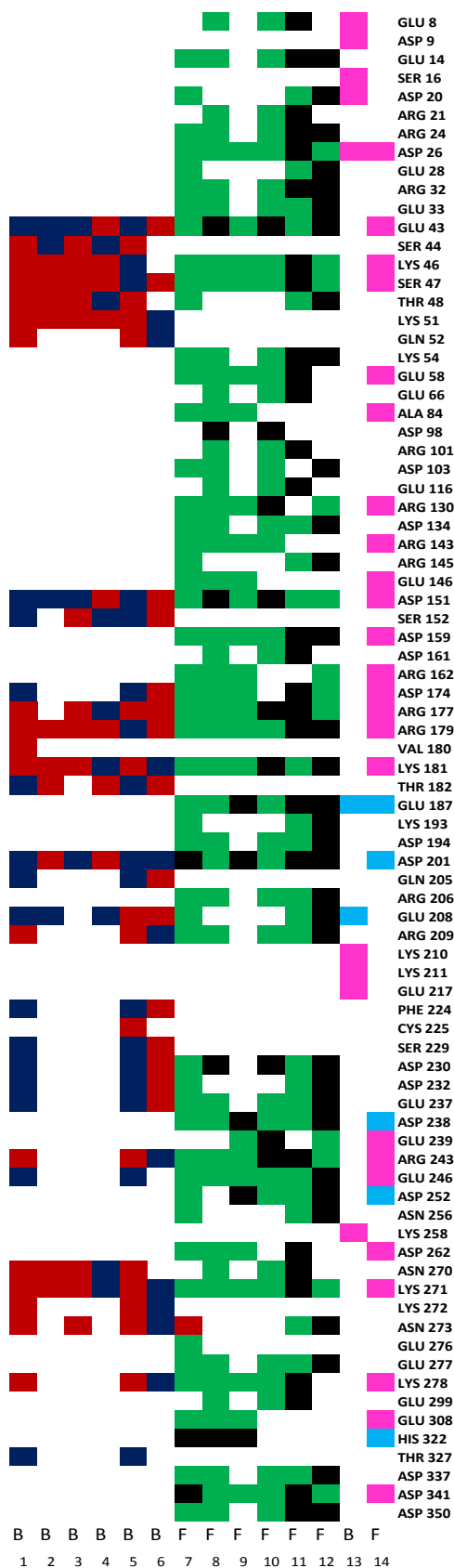


Figure 6.1. Primary structure map for residues that are included using a 1 kcal/mol cutoff for folding and binding of the nucleotide to the α -subunit, and $\alpha\beta$ binding. The color in the block indicates if the residue side chain has a favorable binding energy in red, unfavorable binding energy in blue, favorable folding energy in green, unfavorable folding energy in black, favorable binding energy in magenta and unfavorable binding energy in cyan. Blocks colored white are not considered important and are excluded by the cutoff. The B and F letters in the bottom identifies columns corresponding to binding or folding. The numbers below the letters identify the binding or folding sections of interest in the heterotrimeric G-protein in the following way: 1, $G_{\alpha.GDP}$ binding; 2, $G_{\alpha.GTP.MG}$ binding; 3, $G_{\alpha.GDP.G\beta\gamma}$ binding; 4, binding difference between $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$; 5, binding difference between $G_{\alpha.GDP}$ and $G_{\alpha.GTP.MG}$; 6, binding difference between $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$; 7, $G_{\alpha.GDP}$ folding; 8, $G_{\alpha.GTP.MG}$ folding; 9, $G_{\alpha.GDP.G\beta\gamma}$ folding; 10, folding difference between $G_{\alpha.GTP.MG}$ and $G_{\alpha.GDP.G\beta\gamma}$; 11, folding difference between $G_{\alpha.GDP}$ and $G_{\alpha.GTP.MG}$; 12, folding difference between $G_{\alpha.GDP.G\beta\gamma}$ and $G_{\alpha.GDP}$; 13, $G_{\alpha.GDP.G\beta\gamma}$ binding for the α -subunit; 14, $G_{\alpha.GDP.G\beta\gamma}$ folding for the α -subunit.

relevant components that could be computationally mutated and simulated from the same initial starting structures. These mutations could then be analyzed and compared to the pre-mutated primary structure map. In this way, vast amounts of information can be compiled and compared in simple plots that could occupy as little as one page. Primary structure maps can also be used to display important binding and folding interactions. Figure 6.2 shows how residues important in binding, folding and both in different colors. This figure relates the primary structure maps to tertiary structures in an intuitive way.

Primary structure map information can also be used to display residues according to whether they are energetically favorable or unfavorable. In the case of the $\alpha\beta$ interface, this allows to better identify favorable and unfavorable interaction over a large binding interface. Figure 6.3 shows residues favorable to binding in magenta, and unfavorable to binding in cyan. Possible modifications to the $\alpha\beta$ interface would be easier to suggest from the rendering of the information from primary structure maps to quaternary structure of binding interfaces between proteins. As in the primary structures alone, this way of displaying the quaternary structure would be ideal for comparing results between the native structures and a computationally mutated simulations of G_{ai1} in its three different states.

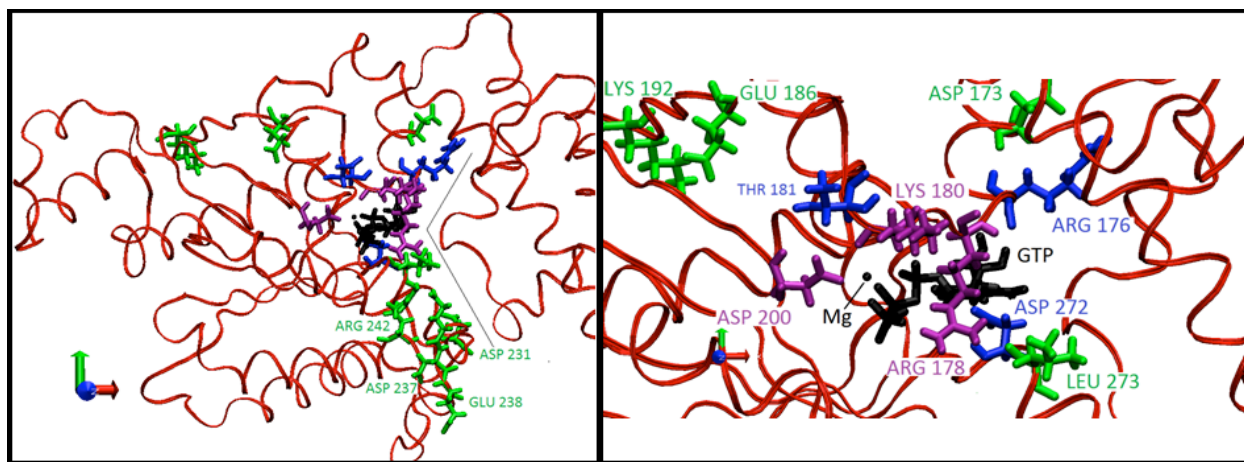


Figure 6.2. Tertiary structure visualization of important folding and binding residues from primary structure map information. Red, α -subunit backbone; black, nucleotide; magenta, important binding residues; green, important folding residues; blue, residues that are both in binding and folding.

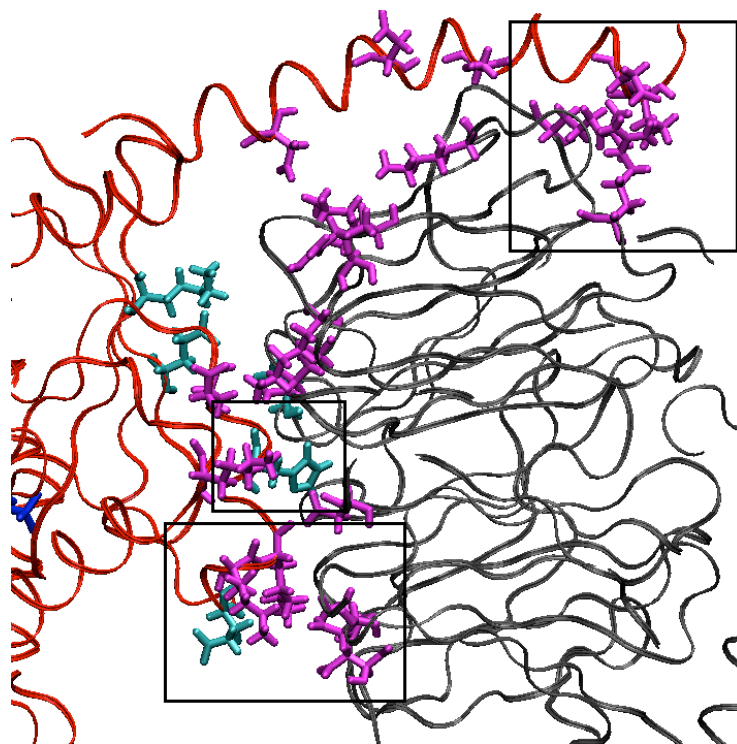


Figure 6.3. Quaternary structure visualization of important $\alpha\beta$ binding residues from primary structure maps. Red, α -subunit backbone; grey, β -subunit backbone; magenta, residues with favorable binding energies; cyan, residues with unfavorable binding energies. The top box shows favorable interactions between α and β subunits at near the N-terminus of the α subunit. Middle box shows HSE 213 in switch region 2 interacting unfavorably with the β -subunit. The bottom box shows GLU 207, LYS 09 and LYS 210 in G_α interaction with ASP 186, ASP 228 and ASP 246 in G_β .

6.2 Future directions

As more structures of variations of heterotrimeric G-proteins are obtained, their simulations and analysis with the method developed for this thesis could be applied for comparison. Structures of the same α -subunit with mutated residues, or other α subunits with large differences in sequences, could give more insight into the mechanisms of signal transduction from the outside environment of the cell and into the cytoplasm.

The analysis presented in this thesis can be applied to new structures of $G_{\alpha i 1}$ in the monomer and complex state, and with variations in its sequences and corresponding structures. These variations of the G-protein and their analysis would add large amounts of structural and energetic information for analysis. Organizing the data into energetic and structural data in succinct plots is important for comparing different types of G-proteins from molecular dynamic

simulations. The structural information could be easily summarized and compared with 2D-RMSDs between different trajectories.

Parameterization of N and C termini post-translational modifications of G-proteins would allow the simulations of this protein in the presence of the cell membrane, and perhaps of a G-protein Coupled Receptor. This type of simulations present challenges and demand long simulations in supercomputers, but the structural and energetic information would help understand the mechanisms of signal transduction in more detail.

The ultimate goal of the accumulation of structural dynamics of these types of proteins is to develop the insight necessary for predict G-protein designs that would behave different from their natural relatives; this goal would require the accumulation of a large set of information that is within current technological capabilities. The compilation, analysis and interpretation of that vast amount of information can be summarized with methods developed in this thesis.

References

1. Casey, P. J. (1995) Protein lipidation in cell signaling, *Science* 268, 221-225.
2. Milligan, G., and Kostenis, E. (2006) Heterotrimeric G-proteins: a short history, *Br J Pharmacol* 147 Suppl 1, S46-55.
3. Downes, G. B., and Gautam, N. (1999) The G-protein subunit gene families, *Genomics* 62, 544-552.
4. Simon, M. I., Strathmann, M. P., and Gautam, N. (1991) Diversity of G-proteins in signal transduction, *Science* 252, 802-808.
5. Neves, S. R., Ram, P. T., and Iyengar, R. (2002) G-protein pathways, *Science* 296, 1636-1639.
6. Dohlman, H. G., and Thorner, J. W. (2001) Regulation of G-protein-initiated signal transduction in yeast: paradigms and principles, *Annu Rev of Biochem* 70, 703-754.
7. Albert, P. R., and Robillard, L. (2002) G-protein specificity: traffic direction required, *Cell Signal* 14, 407-418.
8. Offermanns, S. (2003) G-proteins as transducers in transmembrane signalling, *Prog Biophys Mol Biol* 83, 101-130.
9. Preininger, A. M., Funk, M. A., Oldham, W. M., Meier, S. M., Johnston, C. A., Adhikary, S., Kimple, A. J., Siderovski, D. P., Hamm, H. E., and Iverson, T. M. (2009) Helix dipole movement and conformational variability contribute to allosteric GDP release in G_{αi} subunits, *Biochemistry-Us* 48, 2630-2642.
10. Bunemann, M., Frank, M., and Lohse, M. J. (2003) G_i protein activation in intact cells involves subunit rearrangement rather than dissociation, *Proc Natl Acad Sci USA* 100, 16077-16082.
11. Wang, J., Golebiewska, U., and Scarlata, S. (2009) A self-scaffolding model for G-protein signaling, *J Mol Biol* 387, 92-103.
12. Wang, J., Sengupta, P., Guo, Y., Golebiewska, U., and Scarlata, S. (2009) Evidence for a second, high affinity G_{βγ} binding site on G_{α1} (GDP) subunits, *J Biol Chem* 284, 16906-16913.
13. Kamal, M., Maurice, P., and Jockers, R. (2011) Expanding the concept of G-protein coupled receptor (GPCR) dimer asymmetry towards GPCR-interacting proteins, *Pharmaceuticals* 4, 273-284.
14. Oldham, W. M., and Hamm, H. E. (2006) Structural basis of function in heterotrimeric G-proteins, *Q Rev Biophys* 39, 117-166.

15. Cabrera-Vera, T. M., Vanhauwe, J., Thomas, T. O., Medkova, M., Preininger, A., Mazzoni, M. R., and Hamm, H. E. (2003) Insights into G-protein structure, function, and regulation, *Endocr Rev* 24, 765-781.
16. Mixon, M. B., Lee, E., Coleman, D. E., Berghuis, A. M., Gilman, A. G., and Sprang, S. R. (1995) Tertiary and quaternary structural changes in G_{ia1} induced by GTP hydrolysis, *Science* 270, 954-960.
17. Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G., and Sprang, S. R. (1994) Structures of active conformations of G_{ia1} and the mechanism of GTP hydrolysis, *Science* 265, 1405-1412.
18. Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G., and Sprang, S. R. (1995) The structure of the G-protein heterotrimer G_{ia1} β 1 γ 2, *Cell* 83, 1047-1058.
19. Oldham, W. M., and Hamm, H. E. (2008) Heterotrimeric G-protein activation by G-protein coupled receptors, *Nat Rev Mol Cell Biol* 9, 60-71.
20. Vetter, I. R., and Wittinghofer, A. (2001) The guanine nucleotide-binding switch in three dimensions, *Science* 294, 1299-1304.
21. Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal structure of the β 2 adrenergic receptor-Gs protein complex, *Nature* 477, 549-555.
22. Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules, *Nat Struct Biol* 9, 646-652.
23. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010) Atomic-level characterization of the structural dynamics of proteins, *Science* 330, 341-346.
24. Brooks, B., and Karplus, M. (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, *Proc Natl Acad Sci USA* 80, 6571-6575.
25. Karplus, M., and Petsko, G. A. (1990) Molecular dynamics simulations in biology, *Nat Struct Biol* 347, 631-639.
26. Voelz, V. A., Bowman, G. R., Beauchamp, K., and Pande, V. S. (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39), *J Am Chem Soc* 132, 1526-1528.
27. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proc Natl Acad Sci USA* 103, 915-920.

28. Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review, *Biophys J* 72, 1047-1069.
29. Gilson, M. K., and Honig, B. H. (1987) Calculation of electrostatic potentials in an enzyme active site, *Nature* 330, 84-86.
30. Feig, M., and Brooks, C. L., 3rd. (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations, *Curr Opin Struct Biol* 14, 217-224.
31. Mohan, V., Davis, M. E., McCammon, J., and Pettit, B. M. (1992) Continuum model calculation of solvation free energies: accurate evaluation of electrostatic contributions, *J Phys Chem-Us* 96, 6428-6431.
32. Bashford, D., and Case, D. A. (2000) Generalized-born models of macromolecular solvation effects, *Annu Rev Phys Chem* 51, 129-152.
33. Hendsch, Z. S., and Tidor, B. (1999) Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding, *Protein Sci* 8, 1381-1392.
34. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B., and Jasanoff, A. (2006) Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide, *Biochemistry-Us* 45, 12547-12559.
35. Green, D. F., and Tidor, B. (2005) Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface, *Proteins* 60, 644-657.
36. Lippow, S. M., and Tidor, B. (2007) Progress in computational protein design, *Curr Opin Biotechnol* 18, 305-311.
37. Green, D. F., and Tidor, B. (2003) Evaluation of electrostatic interactions, *Current Protocols in Bioinformatics Chapter 8*, Unit 8 3.
38. Futatsugi, N., and Tsuda, M. (2001) Molecular dynamics simulations of Gly-12-->Val mutant of p21(ras): dynamic inhibition mechanism, *Biophys J* 81, 3483-3488.
39. Gorfe, A. A., Grant, B. J., and McCammon, J. A. (2008) Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins, *Structure* 16, 885-896.
40. Jones, J. C., Temple, B. R., Jones, A. M., and Dohlman, H. G. (2011) Functional reconstitution of an atypical G-protein heterotrimer and regulator of G-protein signaling protein (RGS1) from *Arabidopsis thaliana*, *J Biol Chem* 286, 13143-13150.
41. Raimondi, F., Portella, G., Orozco, M., and Fanelli, F. (2011) Nucleotide binding switches the information flow in ras GTPases, *PLoS Comput Biol* 7, e1001098.

42. Coussens, L. M., Fingleton, B., and Matrisian, L. M. (2002) Matrix metalloproteinase inhibitors and cancer: trials and tribulations, *Science* 295, 2387-2392.
43. Pavlaki, M., and Zucker, S. (2003) Matrix metalloproteinase inhibitors (MMPi): the beginning of phase I or the termination of phase III clinical trials, *Cancer Metastasis Rev* 22, 177-203.
44. Engel, C. K., Pirard, B., Schimanski, S., Kirsch, R., Habermann, J., Klingler, O., Schlotte, V., Weithmann, K. U., and Wendt, K. U. (2005) Structural basis for the highly selective inhibition of MMP-13, *Chem Biol* 12, 181-189.
45. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J Chem Phys* 79, 926-935.
46. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics, *J Am Chem Soc* 112, 6127-6129.
47. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc Chem Res* 33, 889-897.
48. Strockbine, B., and Rizzo, R. C. (2007) Binding of antifusion peptides with HIVgp41 from molecular dynamics simulations: quantitative correlation with experiment, *Proteins: Struct Func Bioinformatics* 67, 630-642.
49. Chachra, R., and Rizzo, R. C. (2008) Origins of resistance conferred by the R292K Neuraminidase mutation via molecular dynamics and free energy calculations, *J Chem Theory Comput* 4, 1526-1540.
50. Rizzo, R. C., Toba, S., and Kuntz, I. D. (2004) A molecular basis for the selectivity of thiazolidine urea inhibitors with stromelysin-1 and gelatinase-A from generalized-Born molecular dynamics simulations, *J Med Chem* 47, 3065-3074.
51. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate calculation of hydration free-energies using macroscopic solvent models, *J. Phys. Chem.* 98, 1978-1988.
52. (2004) AMBER Version 8, University of California at San Francisco, San Francisco, CA.
53. (2007) MOE, Chemical Computing Group, Montreal, Canada.
54. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* 65, 712-725.
55. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J Comput Chem* 25, 1157-1174.

56. Breneman, C. M., and Wiberg, K. B. (1990) Determining atom-centered monopoles from molecular electrostatic potentials - the need for high sampling density in formamide conformational-analysis, *J Comput Chem* 11, 361-373.
57. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B. G., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle, E. S., and Pople, J. A. (1998) Gaussian 98, revision A.9, Gaussian Inc., Pittsburgh PA.
58. Stote, R. H., and Karplus, M. (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation, *Proteins* 23, 12-31.
59. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *J Chem Phys* 98, 10089-10092.
60. Swiss Institute of Bioinformatics (SIB) and European Bioinformatics Institute (EBI), *SWISS-PROT*. <http://www.ebi.ac.uk/swissprot/>.
61. Dingus, J., Wells, C. A., Campbell, L., Cleator, J. H., Robinson, K., and Hildebrandt, J. D. (2005) G-Protein $\beta\gamma$ dimer formation: G_β and G_γ differentially determine efficiency of in vitro dimer formation, *Biochemistry-US* 44, 11882-11890.
62. Pronin, A. N., and Gautam, N. (1992) Interaction between G-protein β and γ subunit types is selective, *Proc Natl Acad Sci USA* 89, 6220-6224.
63. Robillard, L., Ethier, N., Lachance, M., and Hebert, T. E. (2000) $G_{\beta\gamma}$ subunit combinations differentially modulate receptor and effector coupling in vivo, *Cell Signal* 12, 673-682.
64. Schmidt, C. J., Thomas, T. C., Levine, M. A., and Neer, E. J. (1992) Specificity of G-protein beta and γ subunit interactions, *J Biol Chem* 267, 13807-13810.
65. Yan, K., Kalyanaraman, V., and Gautam, N. (1996) Differential ability to form the G-protein $\beta\gamma$ complex among members of the β and γ subunit families, *J Biol Chem* 271, 7141-7146.
66. de Bakker, P. I., Hunenberger, P. H., and McCammon, J. A. (1999) Molecular dynamics simulations of the hyperthermophilic protein sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability, *J Mol Biol* 285, 1811-1830.
67. Kuczera, K., Gao, J., Tidor, B., and Karplus, M. (1990) Free energy of sickling: A simulation analysis, *Proc Natl Acad Sci USA* 87, 8481-8485.

68. Straatsma, T. P., and McCammon, J. A. (1992) Computational alchemy, *Annu Rev Phys Chem* 43, 407-435.
69. Warwicker, J., and Watson, H. C. (1982) Calculation of the electric potential in the active site cleft due to alpha-helix dipoles, *J Mol Biol* 157, 671-679.
70. Gilson, M. K., Sharp, K. A., and Honig, B. H. (1988) Calculating the electrostatic potential of molecules in solution: - Method and error assessment, *J Comp Chem* 9, 327-335.
71. Green, D. F., and Tidor, B. (2004) Escherichia coli glutaminyl-tRNA synthetase is electrostatically optimized for binding of its cognate substrates, *J Mol Biol* 342, 435-452.
72. Kangas, E., and Tidor, B. (2001) Electrostatic complementarity at ligand binding sites: Application to chorismate mutase, *J Phys Chem B* 105, 880-888.
73. Sims, P. A., Wong, C. F., and McCammon, J. A. (2004) Charge optimization of the interface between protein kinases and their ligands, *J Comput Chem* 25, 1416-1429.
74. Archontis, G., Simonson, T., and Karplus, M. (2001) Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase, *J Mol Biol* 306, 307-327.
75. Kuhn, B., and Kollman, P. A. (2000) A ligand that is predicted to bind better to Avidin than Biotin: Insights from computational fluorine scanning, *J Am Chem Soc* 122, 3909-3916.
76. Mardis, K. L., Luo, R., and Gilson, M. K. (2001) Interpreting trends in the binding of cyclic ureas to HIV-1 protease, *J Mol Biol* 309, 507-517.
77. Fujimoto, Y. K., Terbush, R. N., Patsalo, V., and Green, D. F. (2008) Computational models explain the oligosaccharide specificity of cyanovirin-N, *Protein Sci* 17, 2008-2014.
78. Murray, D., McLaughlin, S., and Honig, B. (2001) The role of electrostatic interactions in the regulation of the membrane association of G protein beta gamma heterodimers, *J Biol Chem* 276, 45153-45159.
79. Sarkar, C. A., Lowenhaupt, K., Horan, T., Boone, T. C., Tidor, B., and Lauffenburger, D. A. (2002) Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching", *Nat Biotechnol* 20, 908-913.
80. Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C. (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii, *J Phys Chem A* 101, 3005-3014.
81. Reddy, M. R., Erion, M. D., Agarwal, A., Viswanadhan, V. N., McDonald, D. Q., and Still, W. C. (1998) Solvation free energies calculated using the GB/SA model: Sensitivity of results on charge sets, protocols, and force fields, *J Comput Chem* 19, 769-780.

82. Chocholousova, J., and Feig, M. (2006) Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment, *J Phys Chem B* 110, 17240-17251.
83. Lee, M. S., Salsbury, F. R., and Brooks, C. L. (2002) Novel generalized-Born methods, *J Chem Phys* 116, 10606.
84. Liu, H. Y., Kuntz, I. D., and Zou, X. Q. (2004) Pairwise GB/SA scoring function for structure-based drug design, *J Phys Chem B* 108, 5453-5462.
85. Sorin, E. J., Engelhardt, M. A., Herschlag, D., and Pande, V. S. (2002) RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop, *J Mol Biol* 317, 493-506.
86. Zhu, J., Shi, Y., and Liu, H. (2002) Parametrization of a generalized-Born/solvent-accessible surface area model and applications to the simulations of protein dynamics, *J Phys Chem B* 106, 4844-4853.
87. Chen, J., Im, W., and Brooks, C. L., 3rd. (2006) Balancing solvation and intramolecular interactions: toward a consistent generalized-Born force field, *J Am Chem Soc* 128, 3728-3736.
88. Onufriev, A., Case, D. A., and Bashford, D. (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect, *J Comput Chem* 23, 1297-1304.
89. Hendsch, Z. S., Jonsson, T., Sauer, R. T., and Tidor, B. (1996) Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests, *Biochemistry-U.S.* 35, 7621-7625.
90. Hendsch, Z. S., Nohaile, M. J., Sauer, R. T., and Tidor, B. (2001) Preferential heterodimer formation via undercompensated electrostatic interactions, *J Am Chem Soc* 123, 1264-1265.
91. Nohaile, M. J., Hendsch, Z. S., Tidor, B., and Sauer, R. T. (2001) Altering dimerization specificity by changes in surface electrostatics, *Proc Natl Acad Sci USA* 98, 3109-3114.
92. Lee, L.-P., and Tidor, B. (1997) Optimization of electrostatic binding free energy, *J Chem Phys* 106, 8681.
93. Sulea, T., and Purisima, E. O. (2001) Optimizing ligand charges for maximum binding affinity. A solvated interaction energy approach, *J Phys Chem B* 105, 889-899.
94. Jackson, J. D. Classical Electrodynamics, 3rd Ed., *John Wiley & Sons, Inc. New York*, 1998.
95. Rocchia, W., Alexov, E., and Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions, *J Phys Chem B* 105, 6507-6514.

96. Davis, M. E., and McCammon, J. A. (1989) Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods, *J Comput Chem* 10, 386-391.
97. Holst, M., Baker, N., and Wang, F. (2000) Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples, *J Comput Chem* 21, 1319-1342.
98. Altman, M. D., Bardhan, J. P., White, J. K., and Tidor, B. (2009) Accurate solution of multi-region continuum biomolecule electrostatic problems using the linearized Poisson-Boltzmann equation with curved boundary elements, *J Comput Chem* 30, 132-153.
99. Kangas, E., and Tidor, B. (1998) Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties, *J Chem Phys* 106, 7522.
100. Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P., and Tidor, B. (1998) Computation of electrostatic complements to proteins: a case of charge stabilized binding, *Protein Sci* 7, 206-210.
101. Kangas, E., and Tidor, B. (1999) Charge optimization leads to favorable electrostatic binding free energy, *Phys Rev E* 59, 5958-5961.
102. Lee, L.-P., and Tidor, B. (2001) Barstar is electrostatically optimized for tight binding to barnase, *Nat Struct Biol* 8, 73-76.
103. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci* 3, 211-226.
104. Schaefer, M., and Karplus, M. (1996) A comprehensive analytical treatment of continuum electrostatics, *J Phys Chem-Us* 100, 1578-1599.
105. Im, W., Lee, M. S., and Brooks, C. L., 3rd. (2003) Generalized born model with a simple smoothing function, *J Comput Chem* 24, 1691-1702.
106. Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999) Application of a pairwise generalized-Born model to proteins and nucleic acids: inclusion of salt effects, *Theor Chem Acco: Theor, Comp, Model* 101, 426-434.
107. Research Collaboratory for Structural Bioinformatics (RCSB). Protein Data Bank.
108. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, *J Mol Biol* 285, 1735-1747.
109. Brunger, A. T., and Karplus, M. (1988) Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison, *Proteins* 4, 148-156.
110. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J Comput Chem* 4, 187-217.

111. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J Comput Chem* 26, 1781-1802.
112. MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B* 102, 3586-3616.
113. MacKerell, A. D., Wiórkiewicz-Kuczera, J., and Karplus, M. (1995) An all-atom empirical energy function for the simulation of nucleic acids, *J Am Chem Soc* 117, 11946-11975.
114. Green, D. F., Kangas, E., Hendsch, Z. S., and Tidor, B. (2003) ICE - integrated continuum electrostatics., *MIT*.
115. Altman, M. D., and Tidor, B. (2003) MultigridPBE - Software for computation and display of electrostatic potentials., *MIT*.
116. Nina, M., Beglov, D., and Roux, B. (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations, *J Phys Chem B* 101, 5239-5248.
117. Banavali, N. K., and Roux, B. (2002) Atomic radii for continuum electrostatics calculations on nucleic acids, *J Phys Chem B* 106, 11026-11035.
118. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J Mol Graph* 14, 33-38.
119. Sigel, I. H. (1987) Isomeric equilibria in complexes of adenosine 5'-triphosphate with divalent metal ions, *Eur J Biochemistry* 165, 65-72.
120. Bardhan, J. P. (2008) Interpreting the Coulomb-field approximation for generalized-Born electrostatics using boundary-integral equation theory, *J Chem Phys* 129, 144105.
121. Bardhan, J. P., Knepley, M. G., and Anitescu, M. (2009) Bounding the electrostatic free energies associated with linear continuum models of molecular solvation, *J Chem Phys* 130, 104108.
122. Joughin, B. A., Green, D. F., and Tidor, B. (2005) Action-at-a-distance interactions enhance protein binding affinity, *Protein Sci* 14, 1363-1369.
123. Selzer, T., Albeck, S., and Scheiber, G. (2000) Rational design of faster associating and tighter binding protein complexes, *Nat Struct Biol* 7, 537-541.

124. Wang, J., Sengupta, P., Guo, Y., Golebiewska, U., and Scarlata, S. (2009) Evidence for a second, high affinity Gbetagamma binding site on Galphai1(GDP) subunits, *J Biol Chem* 284, 16906-16913.
125. Willis, B. T., and Pryor, A. W. (1975) *Thermal vibrations in crystallography*, Cambridge University Press London and New York.
126. Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007) Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms, *J Chem Theor Comp* 3, 2312-2334.
127. Karpen, M. E., Tobias, D. J., and Brooks, C. L., 3rd. (1993) Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV, *Biochemistry-Us* 32, 412-420.
128. Satoh, D., Shimizu, K., Nakamura, S., and Terada, T. (2006) Folding free-energy landscape of a 10-residue mini-protein, chignolin, *Febs Lett* 580, 3422-3426.
129. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1978) The Protein Data Bank: a computer-based archival file for macromolecular structures, *Arch Biochem Biophys* 185, 584-591.
130. Martin, A. C. R. Profit is a program that performs fitting of structures by implementing the McLachlan algorithm. www.bioinf.org.uk/software/profit/index.html.
131. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program, *J Comput Chem* 30, 1545-1614.
132. MATLAB. (August 16, 2010) The MathWorks, Inc. Version 7.11.0.584 (R2010b) 64-bit (glnxa64) www.mathworks.com.
133. Van Eps, N., Preininger, A. M., Alexander, N., Kaya, A. I., Meier, S., Meiler, J., Hamm, H. E., and Hubbell, W. L. (2011) Interaction of a G-protein with an activated receptor opens the interdomain interface in the alpha subunit, *Proc Natl Acad Sci USA* 108, 9420-9424.
134. Vidal, M. (2001) A biological atlas of functional maps, *Cell* 104, 333-339.
135. Honig, B., and Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Science* 268, 1144-1149.
136. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H., and Schreiber, G. (2007) The molecular architecture of protein-protein binding sites, *Curr Opin Struct Biol* 17, 67-76.

137. Carrascal, N., and Green, D. F. (2010) Energetic decomposition with the generalized-Born and Poisson–Boltzmann solvent models: Lessons from association of G-protein components, *J Phys Chem B* 114, 5096-5116.
138. Zoete, V., Meuwly, M., and Karplus, M. (2005) Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition, *Proteins* 61, 79-93.
139. Zoete, V., and Michielin, O. (2007) Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex, *Proteins* 67, 1026-1047.

APPENDIX A

Supplementary material for chapter 3: “Energetic decomposition with GB and PB solvent models”

Figure A.1 displays the structural variation in the backbone throughout the simulations. Figure 2 shows the overall correlation between GB and PB for the net binding energy of all three systems. Tables A.1, A.2 and A.3 give the correlation coefficients and best fit lines for each of the individual residues highlighted in the main text. Tables A.4, A.5 and A.6 give the mean desolvation and mutation energies for this same set of residues. Figure A.3 shows the relationship between gbmV-computed binding energies and the same values computed with PB; Figures A.4–A.9 show these data broken down by individual components.

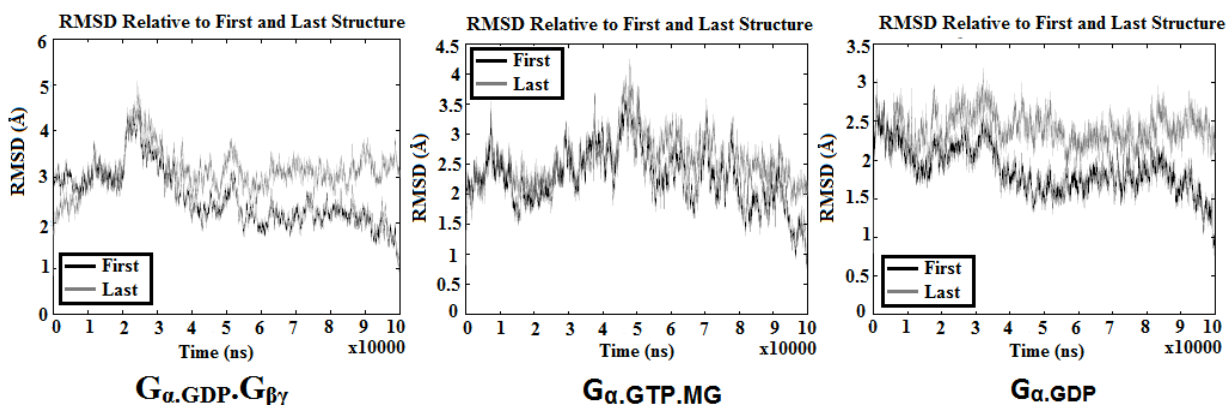


Figure A.1. Structural fluctuations through molecular dynamics simulations. Overall backbone variations (as root-mean-square deviation) from both the first and last frame of the simulation are plotted for each system. All simulations are well-behaved, without any evidence of artifactual conformational changes.

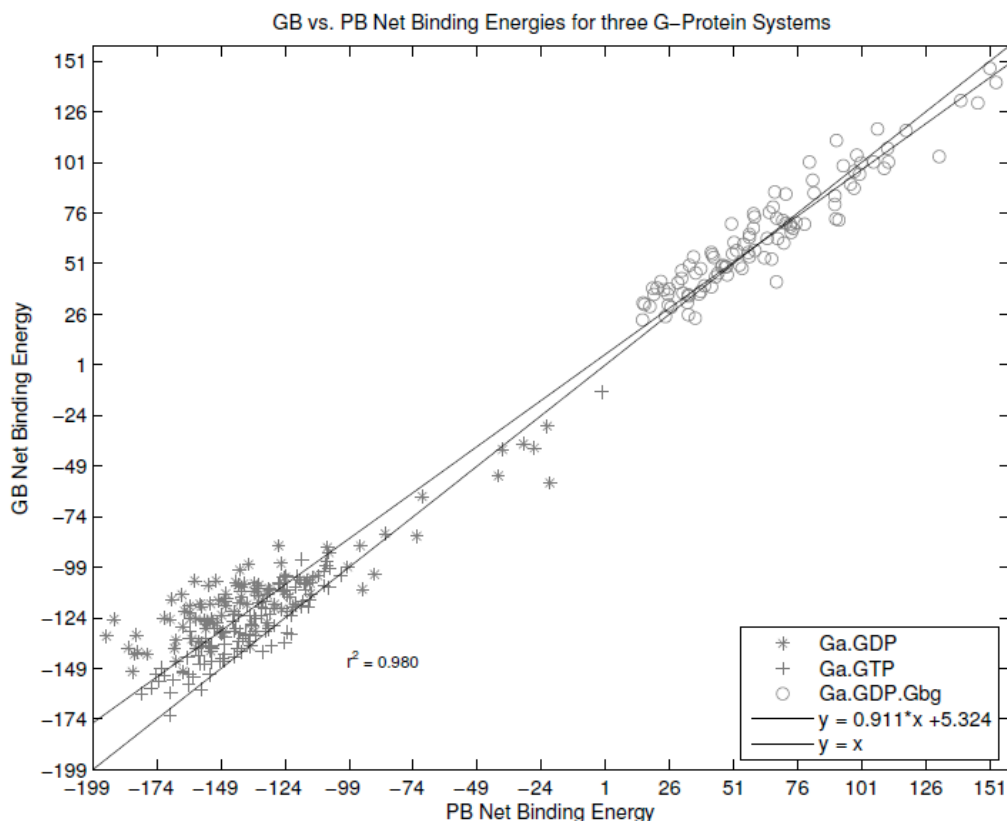


Figure A.2. Variation in single-snapshot net binding energies computed by GB and Pb. The net electrostatic binding free energies computed by both GB and PB are shown for all three systems in a single plot. Over all three systems, the correlation between the two methods is very strong, with near unit slope for the best-fit line. (note: These data are the same as presented in Figure 3.2 of the manuscript.)

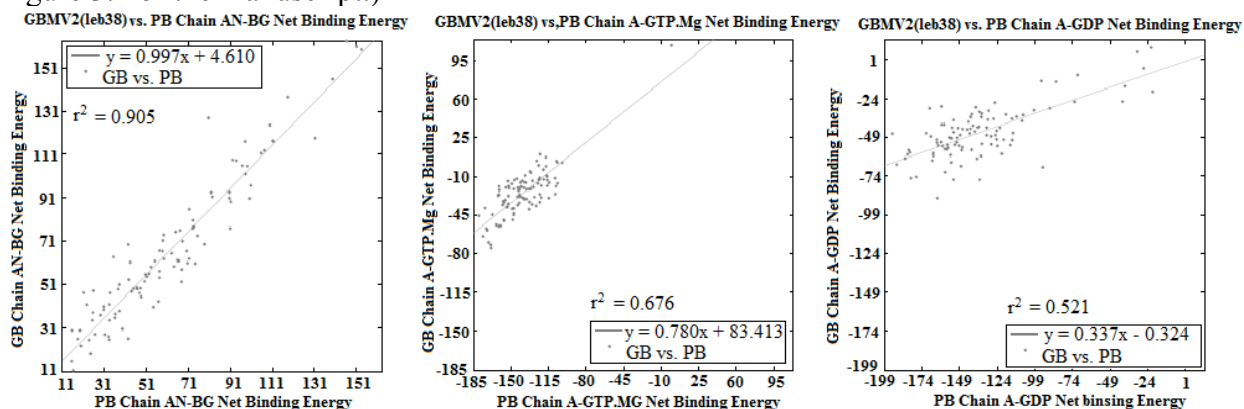


Figure A.3. Variation in single-snapshot net binding energies computed by GBMV and PB. These data can be compared to Figure 3.2 of the text.

Table A.1. $G_{\alpha\text{GDP}}G_{\beta\text{T}}$ residue-specific correlations for select components.^a

Component	Desolvation					Mutation				
	R^2	Best-Fit Line				R^2	Best-Fit Line			
NH	184 α	0.331	0.236	x +	0.12	0.944	1.141	x +	-0.25	
NH	128 β	0.498	0.108	x +	0.011	0.108	0.074	x +	0.015	
NH	145 β	0.453	0.188	x +	0.059	0.746	0.822	x +	-0.004	
NH	144 β	0.334	0.366	x +	0.021	0.721	0.72	x +	-0.339	
CO	90 β	0.67	0.459	x +	0.402	0.988	0.998	x +	-0.707	
CO	117 β	0.273	0.318	x +	0.879	0.84	0.745	x +	-3.05	
CO	143 β	0.27	0.36	x +	0.547	0.854	0.737	x +	-1.811	
CO	89 β	0.543	0.639	x +	-0.186	0.986	0.994	x +	-1.036	
Hsp	54 β	0.176	0.033	x +	0.972	0.239	0.195	x +	-1.446	
Hse	91 β	0.561	0.383	x +	0.721	0.96	1.144	x +	-0.143	
Hse	213 α	0.713	0.446	x +	1.586	0.888	0.975	x +	0.027	
Asp	9 α	0.933	0.88	x +	1.882	0.939	1.072	x +	0.283	
Asp	26 α	0.765	0.727	x +	5.051	0.947	0.942	x +	0.707	
Asp	228 β	0.897	0.745	x +	2.173	0.97	1.116	x +	0.137	
Asp	76 β	0.468	0.131	x +	2.636	0.047	0.04	x +	0.619	
Asn	119 β	0.397	0.309	x +	0.868	0.818	0.65	x +	-1.709	
Asn	22 α	0.921	0.857	x +	0.397	0.959	1.065	x +	-0.038	
Asn	88 β	0.875	0.823	x +	0.98	0.923	1.133	x +	0.195	
Asn	230 β	0.567	0.385	x +	-0.016	0.922	1.11	x +	-1.596	
Ser	16 α	0.304	0.199	x +	2.11	0.925	0.958	x +	-2.366	
Ser	206 α	0.673	0.458	x +	0.707	0.788	0.86	x +	-2.309	
Ser	97 β	0.794	0.31	x +	0.155	0.903	1.119	x +	-0.243	
Glu	186 α	0.761	0.832	x +	0.77	0.906	0.922	x +	-0.635	
Glu	8 α	0.913	0.84	x +	2.996	0.95	0.965	x +	0.956	
Glu	216 α	0.489	0.592	x +	7.268	0.815	0.898	x +	-0.488	
Gln	204 α	0.824	0.529	x +	-0.038	0.735	0.908	x +	-1.215	
Gln	75 β	0.422	0.34	x +	1.009	0.903	0.908	x +	-0.902	
Thr	143 β	0.875	0.571	x +	0.606	0.754	0.763	x +	-0.455	
Thr	182 α	0.672	0.288	x +	1.579	0.881	1.24	x +	-0.651	
Thr	86b	0.587	0.295	x +	0.147	0.664	0.493	x +	-0.038	
Tyr	145 β	0.41	0.208	x +	0.858	0.504	0.47	x +	1.292	
Tyr	59 β	0.815	0.354	x +	0.179	0.731	0.916	x +	-0.093	
Arg	1 α	0.714	0.661	x +	0.413	0.846	0.756	x +	-1.61	
Arg	129 β	0.956	0.816	x +	0.202	0.988	1.055	x +	-0.107	
Arg	68 β	0.942	0.698	x +	0.26	0.987	1.053	x +	-0.186	
Lys	210 α	0.74	0.741	x +	5.683	0.961	0.905	x +	2.897	
Lys	78 β	0.802	0.507	x +	5.994	0.883	0.863	x +	-0.335	
Lys	89 β	0.732	0.844	x +	4.47	0.95	1.084	x +	1.579	
Lys	57 β	0.623	0.446	x +	9.362	0.796	0.859	x +	-0.387	

^aListed components are those highlighted in Figures 3.4 and 3.5 of the text.

Table A.2. $G_{u,GTP,MG}$ residue-specific correlations for select components.^a

Component	Desolvation					Mutation					
	R^2	Best-Fit Line				R^2	Best-Fit Line				
NH	43a	0.199	0.112	x	+	0.111	0.819	1.083	x	+	1.36
NH	45a	0.362	0.199	x	+	0.042	0.791	1.019	x	+	-0.088
NH	48a	0.487	0.209	x	+	-0.006	0.891	1.039	x	+	-0.613
NH	327a	0.457	0.141	x	+	0.023	0.808	1.011	x	+	0.003
CO	176a	0.221	0.207	x	+	0.503	0.842	1.175	x	+	-1.34
CO	175a	0.5	0.386	x	+	-0.059	0.836	0.837	x	+	-3.243
CO	177a	0.751	0.227	x	+	0.065	0.055	0.117	x	+	-0.595
CO	150a	0.578	0.324	x	+	0.03	0.479	0.374	x	+	0.037
Cys	325a	0.523	0.189	x	+	0.011	0.698	0.734	x	+	0.129
Asp	272a	0.963	0.837	x	+	-0.094	0.936	1.395	x	+	-1.585
Asp	200a	0.3	0.38	x	+	3.283	0.872	1.035	x	+	-13.376
Asp	150a	0.792	0.473	x	+	0.501	0.542	0.23	x	+	-0.369
Asn	149a	0.175	0.376	x	+	0.033	0.505	0.167	x	+	0.071
Asn	269a	0.536	0.315	x	+	0.149	0.543	0.939	x	+	0.694
Ser	151a	0.496	0.147	x	+	0.105	0.471	0.428	x	+	-0.535
Ser	47a	0.043	0.131	x	+	0.105	0.975	1.043	x	+	-4.107
Glu	43a	0.273	0.316	x	+	0.882	0.278	0.262	x	+	0.991
Gln	52a	0.015	-0.119	x	+	0.033	0.224	0.287	x	+	0.068
Thr	48a	0.145	0.089	x	+	0.558	0.862	0.886	x	+	-5.348
Thr	327a	0.199	0.218	x	+	0.435	0.729	1.144	x	+	-1.468
Thr	181a	0.001	-0.015	x	+	0.186	0.955	1.027	x	+	-5.205
Tyr	154a	0.766	0.411	x	+	0.067	0.18	0.278	x	+	0.008
Tyr	155a	0.467	0.153	x	+	0.021	0.167	0.427	x	+	0.257
Arg	176a	0.808	0.247	x	+	0.042	0.384	0.189	x	+	-0.04
Arg	178a	0.284	0.554	x	+	2.585	0.801	0.897	x	+	10.546
Lys	51a	0.465	0.113	x	+	1.372	0.29	0.351	x	+	2.124
Lys	180a	0.51	2.31	x	+	0.024	0.579	0.265	x	+	0.089
Lys	270a	0.463	0.429	x	+	1.56	0.623	0.403	x	+	0.513
Lys	46a	0.274	0.244	x	+	8.544	0.497	0.651	x	+	-21.535

^aThis data corresponds to figures 3.7 and 3.8 of the text.

Table A.3. $G_{\alpha, \text{GDP}}$ residue-specific correlations for select components^a.

Component		Desolvation				Mutation			
		R^2	Best-Fit Line			R^2	Best-Fit Line		
NH	43 α	0.144	0.206	x +	0.077	0.906	0.74	x +	-0.62
NH	45 α	0.484	0.212	x +	0.06	0.679	0.862	x +	1
NH	48 α	0.387	0.203	x +	0.016	0.528	0.787	x +	-0.676
NH	327 α	0.508	0.118	x +	0.022	0.387	0.543	x +	0.49
CO	176 α	0.56	0.31	x +	0.123	0.588	0.812	x +	-2.863
CO	175 α	0.433	0.392	x +	-0.038	0.417	0.501	x +	-5.988
CO	177 α	0.65	0.312	x +	0.014	0.017	0.056	x +	-1.21
CO	150 α	0.446	0.242	x +	0.073	0.184	0.351	x +	-1.395
Cys	325 α	0.001	-0.013	x +	0.047	0.347	0.239	x +	0.314
Asp	272 α	0.1	0.131	x +	11.586	0.162	0.214	x +	-24.759
Asp	200 α	0.032	0.358	x +	0.16	0.795	0.347	x +	-0.199
Asp	150 α	0.401	0.492	x +	1.262	0.179	0.17	x +	-0.268
Asn	149 α	0.032	0.05	x +	0.063	0.413	0.15	x +	0.141
Asn	269 α	0.383	0.197	x +	0.182	0.436	0.744	x +	0.642
Ser	151 α	0.406	0.114	x +	0.251	0.619	0.534	x +	0.119
Ser	47 α	0.185	0.114	x +	0.231	0.799	0.684	x +	-3.447
Glu	43 α	0.896	0.68	x +	0.979	0.851	0.358	x +	1.689
Gln	52 α	0	-0.02	x +	0.018	0.085	0.191	x +	-0.481
Thr	48 α	0.481	0.165	x +	0.313	0.779	1.003	x +	-5.227
Thr	327 α	0.007	0.037	x +	0.292	0	0.013	x +	-1.161
Thr	181 α	0.031	0.142	x +	0.017	0.797	0.732	x +	-0.12
Tyr	154 α	0.116	0.128	x +	0.154	0.189	-0.259	x +	0.085
Tyr	155 α	0.006	0.036	x +	0.002	0.005	-0.032	x +	0.425
Arg	176 α	0.036	0.143	x +	1.064	0.107	0.111	x +	-0.423
Arg	178 α	0.511	0.713	x +	-0.53	0.716	0.708	x +	9.001
Lys	51 α	0.359	0.104	x +	1.154	0.346	0.236	x +	0.325
Lys	180 α	0.81	1.033	x +	1.485	0.944	0.797	x +	0.877
Lys	270 α	0.262	0.33	x +	3.671	0.435	0.348	x +	-3.23
Lys	46 α	0.002	-0.017	x +	6.413	0.663	0.451	x +	-30.125

^aThese data corresponds to Figures 3.10 and 3.11 of the text.

Table A.4. $G_{\alpha, \text{GDP}} \cdot G_{\beta}$ residue-specific correlations for select components.^a

Component	Desolvation				Mutation			
	Mean		Std. Dev.		Mean		Std. Dev.	
	PB	GB	PB	GB	PB	GB	PB	GB
NH 184 α	0.573	0.255	0.087	0.036	-2.141	-2.693	1.613	1.894
NH 128 β	0.03	0.015	0.093	0.014	0.076	0.021	0.248	0.056
NH 145 β	0.308	0.117	0.106	0.03	0.154	0.123	0.599	0.57
NH 144 β	0.213	0.099	0.045	0.029	0.161	-0.223	0.741	0.629
CO 90 β	2.431	1.518	0.768	0.431	-3.045	-3.744	4.641	4.659
CO 117 β	2.868	1.792	0.378	0.23	-1.111	-3.878	2.213	1.799
CO 143 β	2.575	1.473	0.498	0.344	0.574	-1.387	2.448	1.954
CO 89 β	2.872	1.65	0.278	0.241	-0.684	-1.716	3.388	3.392
Hsp 54 β	5.572	1.157	6.719	0.533	-2.532	-1.939	3.04	1.214
Hse 91 β	1.469	1.285	1.336	0.684	-0.97	-1.254	3.125	3.651
Hse 213 α	2.077	2.512	0.971	0.513	1.653	1.639	1.979	2.048
Asp 9a	9.274	10.047	8.441	7.695	-2.471	-2.367	5.307	5.874
Asp 26 α	10.256	12.503	4.728	3.927	-8.71	-7.494	9.303	9.002
Asp 228 β	7.585	7.823	6.785	5.338	-9.689	-10.674	9.776	11.076
Asp 76 β	5.54	3.362	4.583	0.877	-3.132	0.493	5.711	1.058
Asn 119 β	4.274	2.189	0.907	0.445	-2.553	-3.367	3.78	2.715
Asn 22 α	0.756	1.045	1.244	1.111	-0.009	-0.047	2.751	2.993
Asn 88 β	2.547	3.076	1.401	1.232	0.157	0.373	2.467	2.91
Asn 230 β	1.871	0.704	1.237	0.632	-0.674	-2.345	3.155	3.65
Ser 16 α	4.304	2.964	1.045	0.376	-5.114	-7.268	3.613	3.599
Ser 206 α	4.715	2.868	1.411	0.788	0.405	-1.961	2.612	2.532
Ser 97 β	1.846	0.728	1.479	0.515	-0.621	-0.938	3.356	3.953
Glu 186 α	16.46	14.462	7.956	7.586	6.055	4.948	8.428	8.163
Glu 8 α	11.731	12.846	8.539	7.502	-11.575	-10.216	10.329	10.23
Glu 216 α	9.977	13.171	2.294	1.94	-12.503	-11.717	2.978	2.962
Gln 204 α	6.291	3.292	2.074	1.209	-3.467	-4.364	2.954	3.13
Gln 75 β	3.606	2.235	1.001	0.523	-1.4	-2.173	2.349	2.244
Thr 143 β	1.686	1.57	1.622	0.991	0.789	0.147	2.528	2.221
Thr 182 α	2.618	2.333	1.421	0.499	-0.524	-1.301	2.259	2.984
Thr 86 β	0.433	0.275	0.377	0.145	-0.478	-0.274	0.838	0.507
Tyr 145 β	3.49	1.585	1.312	0.427	2.144	2.3	1.966	1.301
Tyr 59 β	2.083	0.916	0.671	0.263	-0.803	-0.828	1.338	1.433
Arg 15 α	7.771	5.553	3.858	3.02	-1.763	-2.943	5.19	4.265
Arg 129 β	4.602	3.956	3.162	2.639	-14.454	-15.358	10.581	11.233
Arg 68 β	3.606	2.778	3.213	2.312	-11.048	-11.815	10.216	10.826
Lys 210 α	13.2	15.47	4.9	4.224	-11.907	-7.882	10.331	9.54
Lys 78 β	12.072	12.113	6.907	3.91	-9.631	-8.646	9.64	8.85
Lys 89 β	9.626	12.592	2.807	2.768	-9.428	-8.638	7.224	8.03
Lys 57 β	14.611	15.88	4.327	2.445	-19.547	-17.177	4.383	4.219

^aListed components are those highlighted in Figures 3.4 and 3.5 of the text.

Table A.5. $G_{\alpha, \text{GTP.MG}}$ residue-specific correlations for select components.^a

Component		Desolvation				Mutation			
		Mean		Std. Dev.		Mean		Std. Dev.	
		PB	GB	PB	GB	PB	GB	PB	GB
NH	43 α	0.322	0.147	0.122	0.031	-7.621	-6.896	1.669	1.997
NH	45 α	0.394	0.121	0.1	0.033	-2.719	-2.858	1.66	1.901
NH	48 α	0.188	0.033	0.07	0.021	-5.38	-6.204	1.128	1.242
NH	327 α	0.104	0.038	0.073	0.015	-0.248	-0.248	0.461	0.519
CO	176 α	1.768	0.869	0.411	0.181	-2.013	-3.705	1.169	1.497
CO	175 α	1.962	0.699	0.271	0.148	-2.757	-5.551	1.615	1.478
CO	177 α	0.734	0.231	0.496	0.13	0.375	-0.551	0.974	0.49
CO	150 α	0.108	0.065	0.171	0.073	-0.526	-0.16	0.527	0.285
Cys	325 α	0.155	0.04	0.136	0.036	0.481	0.482	0.651	0.572
Asp	272 α	5.249	4.299	7.656	6.53	-3.47	-6.426	7.764	11.196
Asp	200 α	10.707	7.35	1.842	1.276	-79.499	-95.626	9.995	11.074
Asp	150 α	0.975	0.962	1.915	1.018	3.648	0.472	2.754	0.862
Asn	149 α	0.017	0.039	0.029	0.026	-0.264	0.027	0.534	0.125
Asn	269 α	0.672	0.36	0.36	0.155	-2.285	-1.451	1.188	1.513
Ser	151 α	2.134	0.419	1.284	0.269	1.553	0.13	1.411	0.88
Ser	47 α	0.95	0.23	0.204	0.13	-26.662	-31.92	6.014	6.354
Glu	43 α	0.729	1.112	0.947	0.574	8.273	3.16	2.239	1.112
Gln	52 α	0.028	0.029	0.011	0.01	-1.079	-0.242	0.333	0.202
Thr	48 α	4.44	0.954	0.545	0.128	-10.476	-14.627	1.963	1.873
Thr	327 α	1.887	0.846	0.571	0.278	0.234	-1.2	1.194	1.6
Thr	181 α	1.131	0.168	0.227	0.094	-37.057	-43.27	5.093	5.354
Tyr	154 α	0.086	0.102	0.226	0.106	0.042	0.019	0.246	0.161
Tyr	155 α	0.15	0.044	0.17	0.038	-0.134	0.199	0.324	0.339
Arg	176 α	0.808	0.242	1.658	0.456	-2.336	-0.482	1.297	0.396
Arg	178 α	9.076	7.615	1.005	1.045	-67.515	-50.026	6.576	6.591
Lys	51 α	8.621	2.344	3.63	0.6	-19.187	-4.614	3.125	2.037
Lys	180 α	0.021	0.072	0.021	0.067	-2.704	-0.628	1.078	0.376
Lys	270 α	9.558	5.662	3.968	2.501	-14.336	-5.267	6.979	3.564
Lys	46 α	16.336	12.532	1.969	0.919	-108.408	-92.128	5.057	4.673

^aThese data corresponds to Figures 3.10 and 3.11 of the text.

Table A.6. $G_{\alpha, \text{GDP}}$ residue-specific correlations for select components.^a

Component	Desolvation				Mutation			
	Mean		Std. Dev.		Mean		Std. Dev.	
	PB	GB	PB	GB	PB	GB	PB	GB
NH 43a	0.122	0.102	0.056	0.03	-7.688	-6.312	4.432	3.446
NH 45a	0.287	0.12	0.105	0.032	-10.283	-7.868	2.481	2.596
NH 48a	0.164	0.049	0.058	0.019	-8.321	-7.228	1.183	1.283
NH 327a	0.14	0.039	0.096	0.016	-0.263	0.347	0.515	0.449
CO 176a	1.237	0.506	0.319	0.132	-0.519	-3.284	1.643	1.739
CO 175a	1.801	0.669	0.188	0.112	0.147	-5.914	1.82	1.413
CO 177a	1.216	0.393	0.429	0.166	2.364	-1.077	1.608	0.685
CO 150a	0.568	0.211	0.385	0.14	-2.082	-2.126	1.007	0.824
Cys 325a	0.11	0.045	0.075	0.027	0.451	0.422	1.512	0.613
Asp 272a	19.278	14.109	3.318	1.371	-1.192	-25.014	8.855	4.71
Asp 200a	0.021	0.168	0.051	0.101	5.567	1.733	3.258	1.268
Asp 150a	2.359	2.423	1.318	1.024	14.294	2.168	4.324	1.743
Asn 149a	0.082	0.068	0.086	0.024	1.793	0.41	1.18	0.275
Asn 269a	0.46	0.273	0.371	0.118	-4.734	-2.881	1.439	1.622
Ser 151a	1.91	0.469	0.841	0.151	3.886	2.194	1.865	1.265
Ser 47a	0.586	0.298	0.568	0.15	-22.346	-18.729	5.509	4.214
Glu 43a	2.128	2.426	3.4	2.443	16.995	7.778	12.213	4.743
Gln 52a	0.027	0.018	0.009	0.009	-2.83	-1.023	0.532	0.348
Thr 48a	4.281	1.021	0.602	0.144	-11.171	-16.43	1.914	2.175
Thr 327a	1.713	0.356	0.48	0.213	2.541	-1.127	1.446	1.362
Thr 181a	0.009	0.019	0.021	0.017	1.648	1.087	1.571	1.289
Tyr 154a	0.214	0.181	0.24	0.09	-0.877	0.312	1.183	0.705
Tyr 155a	0.017	0.003	0.026	0.012	0.291	0.416	0.615	0.291
Arg 176a	4.688	1.736	1.068	0.808	-13.624	-1.934	4.379	1.485
Arg 178a	2.854	1.504	1.19	1.187	-19.457	-4.768	8.019	6.707
Lys 51a	10.299	2.23	3.586	0.625	-34.573	-7.828	7.674	3.077
Lys 180a	10.084	11.898	3.693	4.238	-89.848	-70.732	28.474	23.352
Lys 270a	9.455	6.791	2.404	1.55	-41.315	-17.617	7.256	3.831
Lys 46a	2.515	6.371	2.054	0.762	-83.241	-67.692	12.088	6.7

^aThese data corresponds to Figures 3.10 and 3.11 of the text.

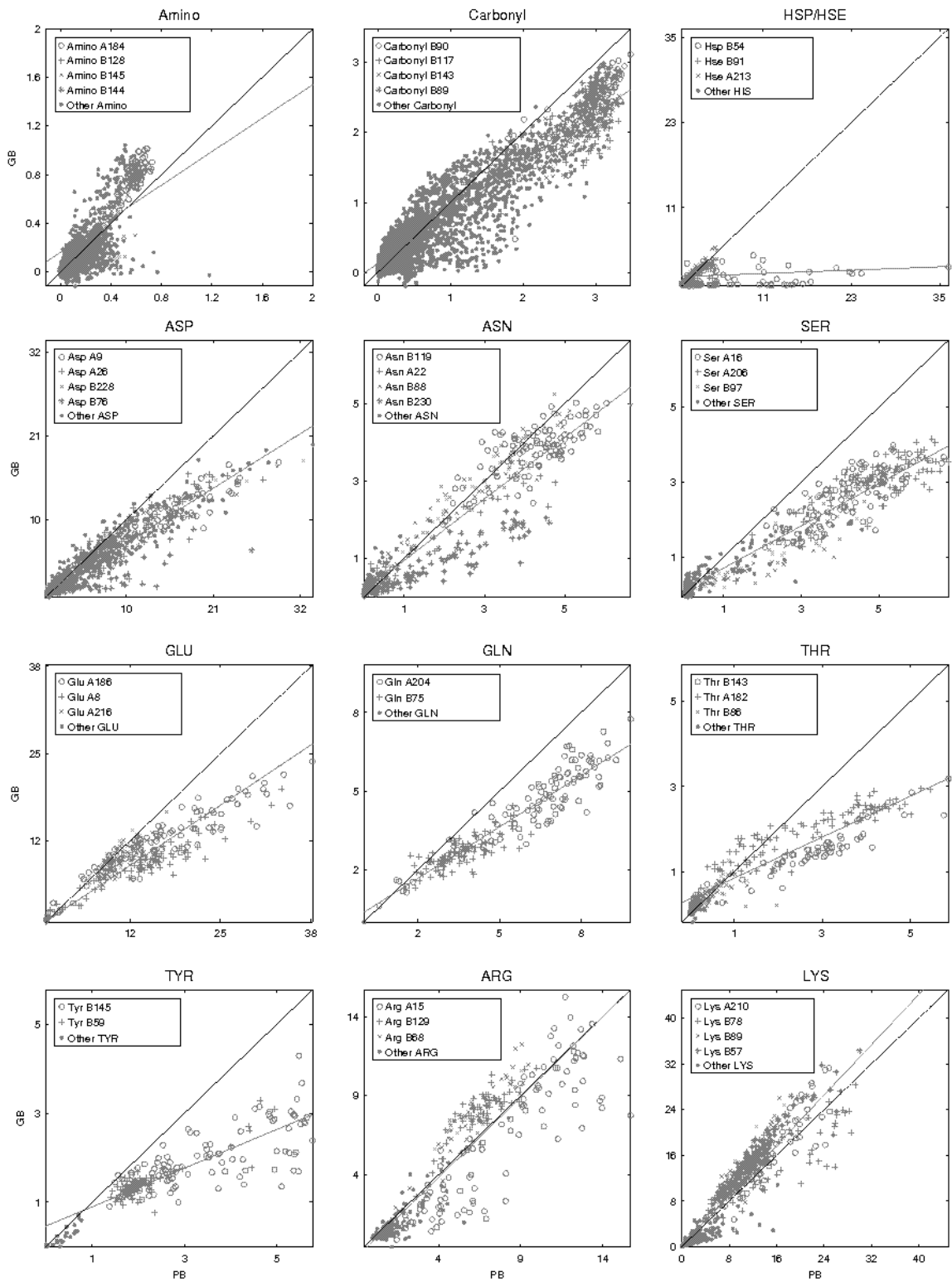


Figure A.4. $G_{\alpha, \text{GDP}} \cdot G_{\beta\gamma}$ group desolvation penalties, by type, using GBMV model. These data can be compared to Figure 3.4 of the text.

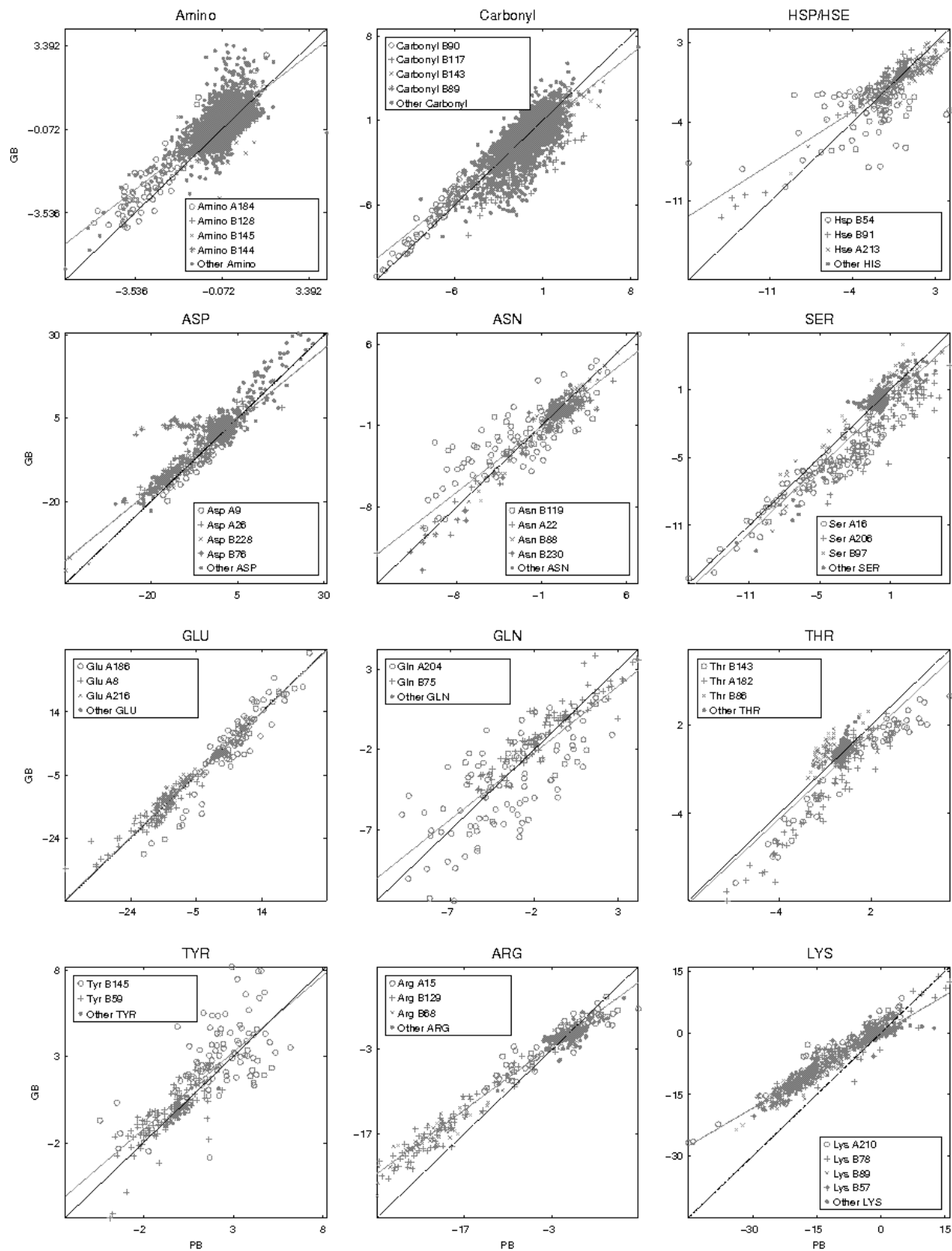


Figure A.5. $G_{\alpha, GDP} \cdot G_{\beta\gamma}$ group mutation penalties, by type, using GBMV model. These data can be compared to Figure 3.5 of the text.

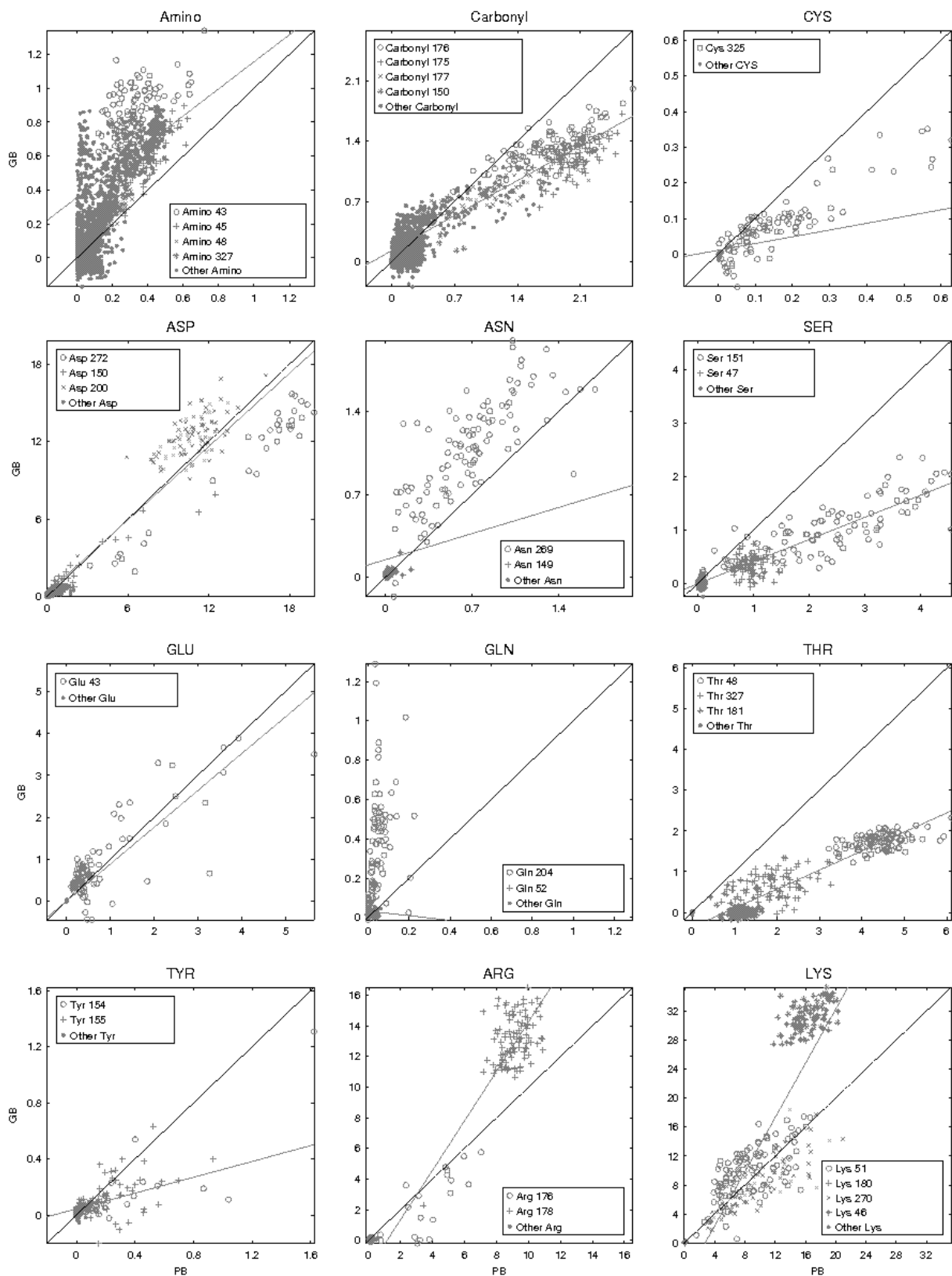


Figure A.6. $G_{a,GTP,MG}$ group desolvation penalties, by type, using GBMV model. These data can be compared to Figure 3.7 of the text.

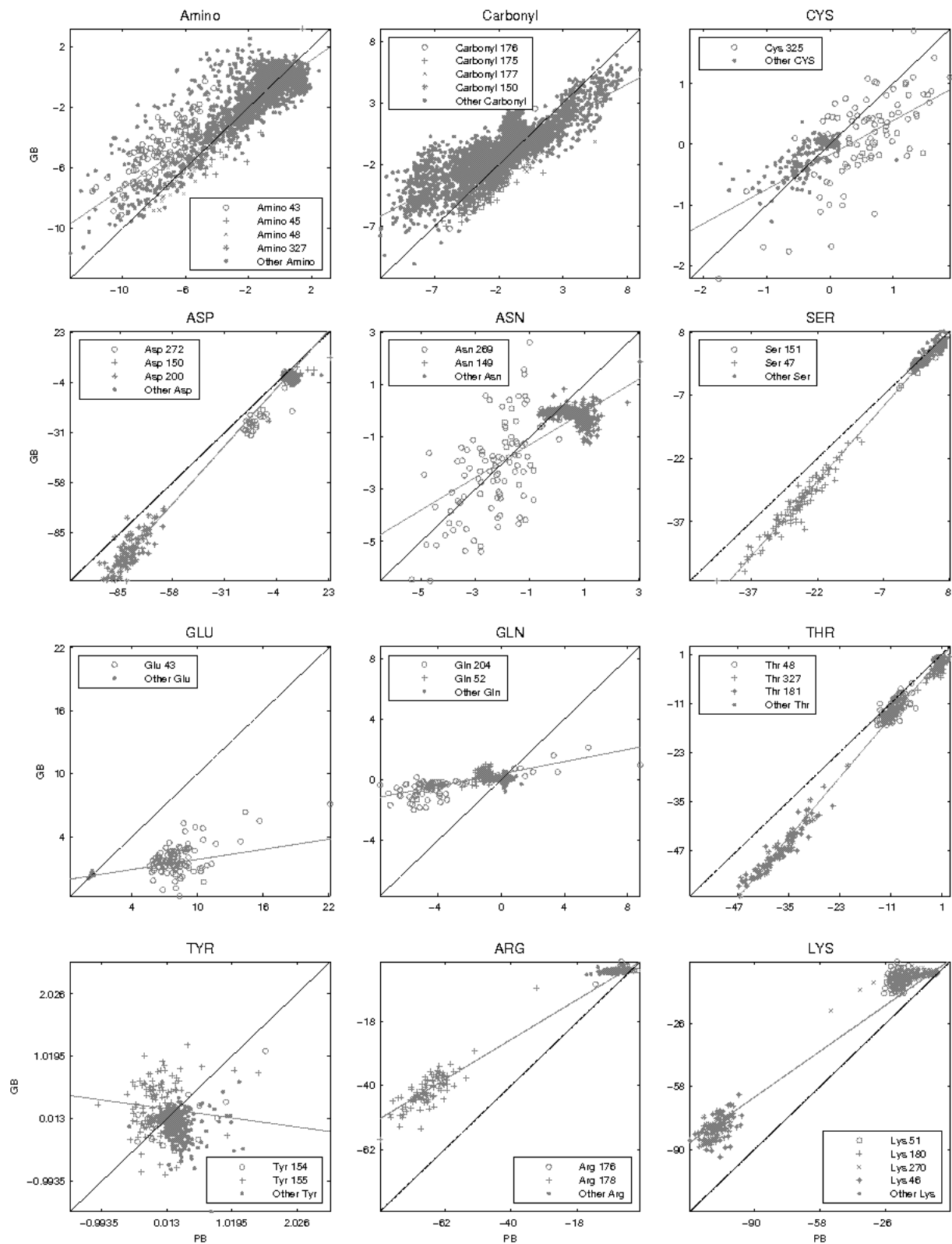


Figure A.7. $G_{a.GTP.MG}$ group mutation penalties, by type, using GBMV model. These data can be compared to Figure 3.8 of the text.

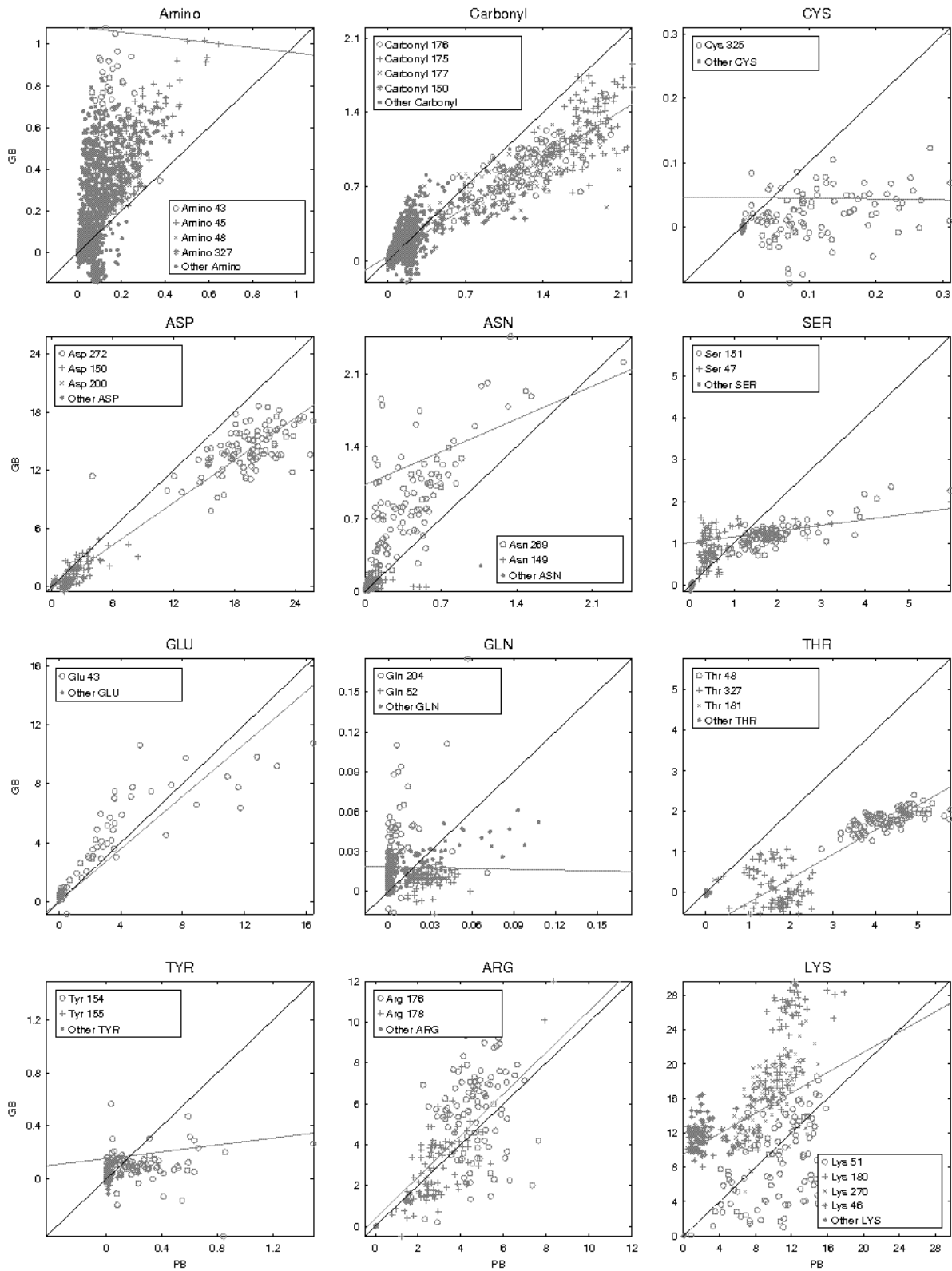


Figure A.8. $G_{\alpha,GDP}$ group desolvation penalties, by type, using GBMV model. These data can be compared to Figure 3.10 of the text.

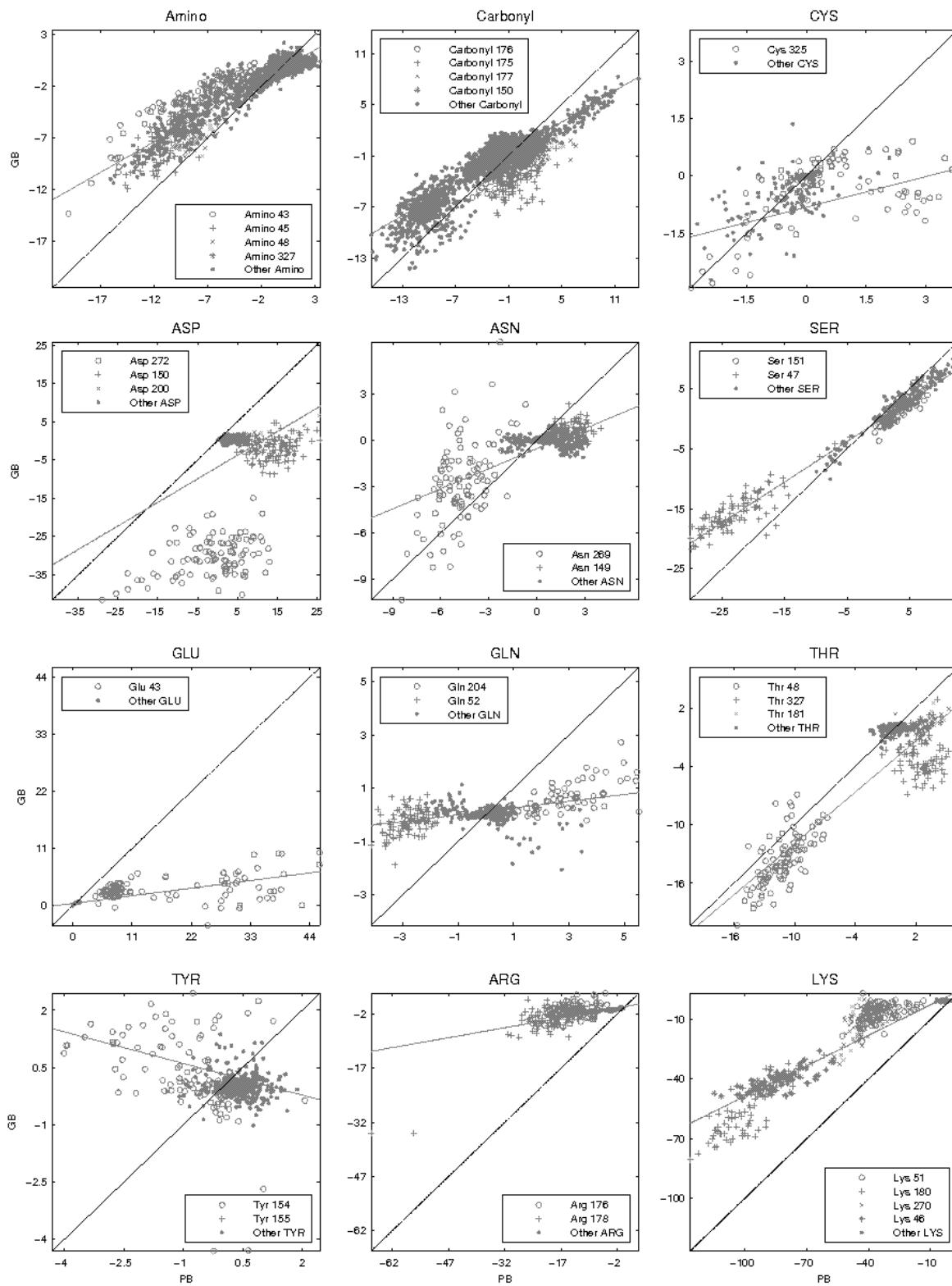


Figure A.9. $G_{\alpha, \text{GDP}}$ group mutation penalties, by type, using GBMV model. These data can be compared to Figure 3.11 of the text.

APPENDIX B

Completing missing structural information in the crystallographic structures from homology modeling between the three systems

Crystal structures used as a starting point for simulations do not have structural information for all the residues in their sequence because of lack of x-ray diffraction. Complete structures were built by combining information from all the structures, Missing Structural information was filled by simply aligning the common available residues around the missing structure with that from a fully resolved structure. A simple least-square alignment of the α -carbons of the back bone was sufficient to merge the structures. The program PROFIT was used for aligning the structures to α -carbons that were present in both structures. The regions were added to the PDBs with missing structural information by copying and pasting the text corresponding lines of text from one PDB file to the other. The following color coded sequence alignments describes what parts were missing, added and not included in the three structures:

```
1GP2_A  GCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEAGY
1GIA_A  GCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEAGY
1GDD_A  GCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEAGY

1GP2_A  SEEEEECKQYKAVVYSNTIQSIIAIIIRAMGRLKIDFGDAARADDARQLFVLAGAAEEGFMTA
1GIA_A  SEEEEECKQYKAVVYSNTIQSIIAIIIRAMGRLKIDFGDAARADDARQLFVLAGAAEEGFMTA
1GDD_A  SEEEEECKQYKAVVYSNTIQSIIAIIIRAMGRLKIDFGDAARADDARQLFVLAGAAEEGFMTA

1GP2_A  ELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYLNDLDRIAQPNYIPTQQDVLRTRVKT
1GIA_A  ELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYLNDLDRIAQPNYIPTQQDVLRTRVKT
1GDD_A  ELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYLNDLDRIAQPNYIPTQQDVLRTRVKT

1GP2_A  TGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEMN
1GIA_A  TGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEMN
1GDD_A  TGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEMN

1GP2_A  RMHESMKLFDSICNNKWFTDTSIIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEAAA
1GIA_A  RMHESMKLFDSICNNKWFTDTSIIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEAAA
1GDD_A  RMHESMKLFDSICNNKWFTDTSIIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEAAA

1GP2_A  YIQCFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF
1GIA_A  YIQCFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF
1GDD_A  YIQCFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF
```

- █ Sequence with structural information from crystal structure.
- █ Structure missing in crystal but added from 1GIA.
- █ Structure missing in crystals but added from 1GDD.
- █ Structure missing in crystals, not included in simulations.
- █ Structure missing in crystal but added from 1GP2.