Journal of Theoretical and Computational Chemistry
1930001 (17 pages)

DOI: 10.1142/S0219633619300015

World Scientific
www.worldscientific.com

Computational chemistry methods to investigate the effects caused by DNA variants linked with disease

Mahesh Koirala and Emil Alexov*

Department of Physics and Astronomy
Clemson University, Clemson, SC 29630, USA
**ealexov@clemson.edu*

Received 5 August 2019
Accepted 3 September 2019
Published 4 October 2019

Computational chemistry offers variety of tools to study properties of biological macromolecules. These tools vary in terms of levels of details from quantum mechanical treatment to numerous macroscopic approaches. Here, we provide a review of computational chemistry algorithms and tools for modeling the effects of genetic variations and their association with diseases. Particular emphasis is given on modeling the effects of missense mutations on stability, conformational dynamics, binding, hydrogen bond network, salt bridges, and pH-dependent properties of the corresponding macromolecules. It is outlined that the disease may be caused by alteration of one or several of above-mentioned biophysical characteristics, and a successful prediction of pathogenicity requires detailed analysis of how the alterations affect the function of involved macromolecules. The review provides a short list of most commonly used algorithms to predict the molecular effects of mutations as well.

Keywords: Computational chemistry; mutations; diseases; stability; binding; human DNA variants; pathogenic mutations.

1. Overview of Computational Chemistry Approaches

Computational chemistry is a sub-field of theoretical chemistry where the main focus is solving chemically related problems by calculations.[1] It simply uses the mathematical algorithms, statistics and large databases to integrate chemical theory and modeling. Frequently, it utilizes high performance computing methods to solve problems that require either huge amount of data or extensive calculations. When computational chemistry deals with molecular biology phenomena, typically one computes variety of properties as folding and binding free energies, electrostatic potential, vibrational frequencies and normal modes, electronic excitation energy,

*Corresponding author.

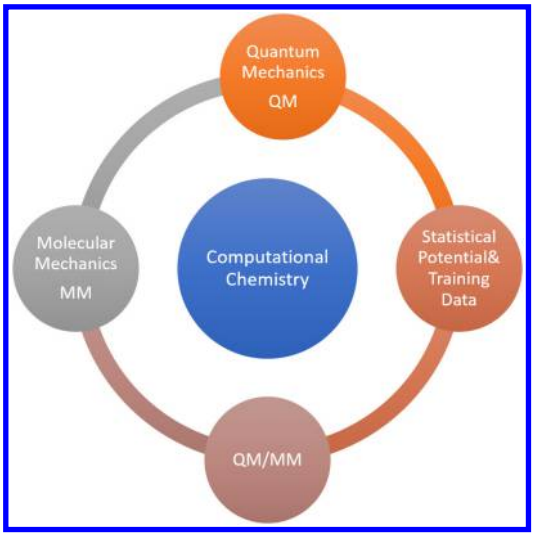

Fig. 1. Different approaches used in computational chemistry to model biological systems.

NMR chemical shifts and coupling constants, reaction path and reaction rate.[2,3] These properties are investigated via various modeling techniques, which range from *ab-initio*, semi-empirical and molecular mechanics (MM).[4,5] Here, we review different approaches of computational chemistry depending on their level of detail from quantum mechanics (QM), molecular dynamics (MD) to macroscopic approaches as molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) and statistical methods (Fig. 1).

Quantum mechanical approach provides the greatest level of details and is used when one is concerned about finest details of the phenomena being investigated. This requires solving the Schrodinger equation for the system of interest and further obtaining relevant energies and properties of molecules.[6,7] Various techniques are applied and here we briefly mention some of them. The Hartree–Fock (HF)[8] method is one of the commonly used approaches in QM modeling of biological macromolecules. An important feature of the HF method is the mean field approximation. Another alternative method in QM for macromolecular modeling is the Density Functional Theory (DFT).[9] In the DFT total energy is expressed in terms of total energy density rather than wave function. The Quantum Monte Carlo (QMC)[10] which uses various energy functions as variational, diffusion and Green's functions and then applies Monte Carlo integration is an another alternative. While QM modeling provides the most physical insights, it is computationally prohibited to be applied on large systems as biological macromolecules, thus one takes advantage of hybrid approaches.[11–15] Indeed, combination of QM and MM (QM/MM) is frequently used in computational chemistry. Thus, QM describes the chemical processes that are localized in space and time and those phenomena that involve slow motion of atoms during a reaction are described via MM.[16]

Other common approaches in computational chemistry are the first-principle approaches that include MD and Monte Carlo (MC) simulations. The MD simulations model the time-dependent behavior of a system of interest within the framework of classical mechanics.[17–20] It is typically used to probe the conformational flexibility of macromolecules, to deliver snapshot of structural conformations, to guide binding via steered MD, to fold macromolecules and many other applications.[21,22] The MC methods, on the other hand, are used to sample energy landscape and to deliver the probability for each accessible conformation.[22,23] Both MD and MC could be quite computationally demanding, if the simulated system is large and requires either long simulation time or sampling of many different states. Thus, one applies techniques that accelerate the sampling and improve the convergence.[14,13,24]

The trajectories generated via MD simulations can be used to compute the energy of the system of interest. This approach is known as molecular mechanics Poisson–Boltzmann (or Generalized Born) surface area (MM-PB/GBSA) method. Thus, using the conformation generated with MD simulations, one partitions the energy into vacuum (MM), polar solvation (PB or GB) and nonpolar (SA) components.[12,25,26]

Approaches based on statistical potentials are frequently used in computational chemistry if one wants to investigate a particular class of problems. Thus, if one is concerned about protein folding; binding and aggregation; or in general of protein structure prediction and in fold recognition,[27,28] instead of using first-principle approaches, the corresponding potentials can be delivered from statistical analysis. In such a case, knowledge-based potentials (KBP)[11,15,29] are delivered by analysis of existing protein structures in Protein Data Bank (PDB).[30]

2. Overview of Computational Chemistry Methods to Predict Changes Caused by Mutations

In this section, we will review different algorithms of computational chemistry that are used to model the effect of mutations on protein wild type properties. We refer to protein wild type properties as folding and binding free energies, conformational dynamics, hydrogen bonds, salt bridges, electrostatic forces, and pH-dependence. We focus on these properties because they can be modeled via the computational chemistry approaches outlined above. Furthermore, since the goal of this review is to review the computational chemistry approaches with regard to predicting disease-causing mutations (variants), it should be pointed out that it was demonstrated that there is a linkage between the change of the above-mentioned properties and propensity given mutation to be pathogenic (for more details see Refs. 31–34). Thus, the ability to correctly predict the changes of wild type properties of the corresponding macromolecules caused by mutations is critical for disease diagnostics.

2.1. *Prediction of stability change*

Biological macromolecules function by adopting a particular 3D structure. A mutation, a change of the amino acid or nucleic acid, can affect the ability of the

Table 1. Different approaches for predicting stability change upon mutation.

Type	Name	Description	Reference
FEP & TI	MD simulations	Uses alchemical transformation from wild type to mutant protein to evaluate the stability change upon mutation.	38–42
MM-PB/GASA	PoPMuSiC	Uses combination of statistical potentials and neutral method to estimate ΔΔG upon mutation.	43
	SAAFEC	Predict the effect of single point mutation on protein folding free energy using a knowledge based MM/PBSA approach.	44
	HoTMuSiC	Uses artificial neutral and solvent accessibility dependent combination of statistical potential to predict stability change upon single mutation.	45
Empirical & machine learning	DUET	Uses computational approach estimate ΔΔG upon mutations in protein using support vector machine.	46
	FoldX	Predicts ΔΔG upon mutation using empirical force field.	47
	I-Mutant	Predicts ΔΔG upon mutation through support vector machine.	48
	SDM	Uses statistical potential energy function to estimate ΔΔG upon mutation.	49
	CUPSAT	Uses mean force atom pair and torsion angle potentials to evaluate ΔΔG upon mutation.	50
	AUTO-MUTE	Uses machine learning methods and knowledge based statistical contact potential to calculate ΔΔG upon mutation.	51

macromolecule to fold properly. This is addressed by computing the effect of mutations on folding free energy. In what follows, we outline some of the most popular approaches to predict the change of the folding free energy caused by mutations (see also Refs. 32, 35–37). The approaches can be grouped into first-principle (free energy perturbation (FEP) of thermodynamic integration (TI)), MM-PB/GASA and empirical methods (Table 1).

The FEP/TI protocols are used in computational chemistry for computing free energy difference between wild type and mutant proteins.[38–42] While they are considered to be the most accurate, the FEP/TI are computationally demanding and cannot be applied on large-scale modeling. Reasonable alternative are the methods based on MMPB/GBSA approach. Several such approaches were developed and are available via webservers (Table 1). The short list includes PoPMuSiC,[43] SAAFEC[44] and others (Table 1). Since MMPB/GBSA is known to overestimate the energy changes, most of the protocols use adjustable parameters to provide optimal weight of the corresponding energy components.[44] The optimization of the parameters is done by benchmarking against experimental data of folding free energy changes caused by mutations (taken from various databases as ProTherm,[45,52] ProDataTherm[53] and SPROUTS[54]). Empirical or statistical methods use either empirical formula or machine learning to optimize their predictions using the above databases.

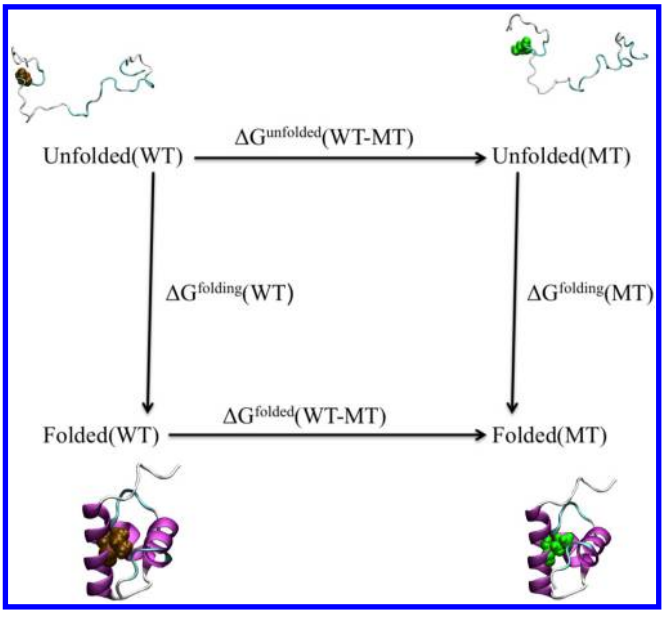

Fig. 2. Thermodynamics cycle for the estimation of folding free energy change upon mutation. The wild type amino acid is shown with brown color, while the mutated with green. Folded state is represented by the corresponding PDB structure (in ribbon presentation) and unfolded state is shown as disordered peptide.

Short list includes DUET,[46] FoldX,[47] I-Mutant,[48] SDM,[49] CUPSAT,[50] and AUTO-MUTE.[51] While these methods are quite fast, they are typically not performing well on cases that are not similar to the cases in the training database.

All above approaches suffer the same uncertainty — the model of unfolded state. It should be emphasized that computing the change of the folding free energy requires either the change of the folding free energy of wild type and mutant protein to be modeled (vertical arrows in Fig. 2) or the change of the free energy of unfolded and folded state upon mutation of the wild type protein (horizontal arrows in Fig. 2). Typically, the above-mentioned approaches either ignore the unfolded state (via training or another procedure) or adopt simplified model of the unfolded state made of one (PoPMuSiC) or several residues (SAAFEC) only. This is clear oversimplification, since it is well documented that unfolded state is not an extended chain of amino acid, instead it is quite compact and there are residual interactions. To address this issue, recently we proposed that unfolded state could be modeled via tools that generate structures for intrinsically disordered proteins.[55]

2.2. *Prediction of binding free energy change*

Similar classification (as for the folding free energy) for the tools for predicting binding free energy can be applied here. The FEP and TI were used to predict the change of the binding free energy caused by mutations.[33,38] However, as mentioned

Table 2. Different approaches for prediction of binding energy change upon mutation.

Type	Name	Description	Reference
FEP & TI	MD simulations	Uses alchemical transformation from wild type to mutant protein to evaluate the affinity change upon mutation.	33, 38
MM-PB/GBSA	BeATMuSiC	Estimates change in binding energy change upon mutation using coarse-grained model and statistical potentials.	56
	MutaBind	Uses MM force fields energies and fast side-chain optimization algorithm to compute the change in binding affinity upon mutation.	57
	SAAMBE	Models the effects of mutations on protein–protein binding free energy using modified MM/PBSA approach and additional set of knowledge-based terms.	58
	SAMPDI	Estimates the effects of mutations on protein–DNA binding affinity using modified MM/PBSA approach and additional set of knowledge-based terms.	59
	PremPDI	Estimate the effect of single mutations on protein–DNA binding affinity using MM force field energies and fast side chain optimization algorithm.	35
Empirical & machine learning	FoldX	Models mutations using rotamer approach and empirical force field.	48
	BindProfX	Estimates the binding free energy change as the logarithm of relative probability of mutant amino acid over wild type ones. It combines interface profile (conservation) with FoldX prediction.	61
	mCSM	Estimates the effect of mutation on protein based on atom distance patterns with Gaussian process regression.	62
	ELASPIC	Machine learning approach to estimate the effects of mutations on protein–protein interaction. Uses sequence, energy and molecular features with decision tree.	63

above, they are quite computationally demanding and cannot be applied on genome-scale investigations. Much more frequently used are MM-PB/GBSA-based approaches (Table 2). The short list includes BeAtMuSiC,[56] MutaBind,[57] SAAMBE,[58] SAMPDI,[59] and PremPDI.[35] They are all trained against databases of experimentally measured changes of the binding free energy as SKEMPI,[32] PBDBind[34,60] and others. Empirical and machine learning algorithms include FoldX,[48] BindProfX,[61] mCSM[62] and ELASPIC[63] (Table 2). Note that in this case, the issue with unfolded state is no longer present. This is because the unfolded state is the same for the complex and separated monomers, so it cancels out in the thermodynamic cycle.

2.3. *Change of conformation dynamics*

Another important characteristic of biological macromolecules is their conformational dynamics. A change of the wild type dynamics can severely affect the function. Typically, one investigates macromolecular dynamics via MD approaches. Most commonly used packages are NAMD,[64] CHARMM,[65] Amber,[66] GROMOS,[67] GROMACS,[68] TINKER,[69] and others. The change of conformational dynamics is

attained computationally by comparing the root mean square deviation (RMSD) of wild type and mutant structures obtained via MD simulations. In addition, frequently researchers use root mean square fluctuations (RMSF) to identify structural regions affected by the mutations.

2.4. *Prediction of change of H–bond network*

Hydrogen bonds are essential for macromolecular stability, flexibility and details of the corresponding reaction. Any deviation of the native H–bond network is expected to affect the function of the corresponding macromolecule. The simplest approach of predicting the effect of mutation on H–bonds is to compare the H–bonds of wild type and mutant structures being obtained either experimentally or computationally (via side chain replacement). In the last case, one can employ energy minimization to reduce plausible conformational clashes and to further optimize the H–bonds before making the assessment. More sophisticated approaches require MD simulations and analysis of the corresponding snapshots. All MD packages listed above provide tools for such kind of analysis, the most used is VMD.[70] Some of other tools include Chimera,[71] PyMOL[71] and RASMOL.[72]

2.5. *Prediction of change of salt-bridges*

Salt bridges are also essential component of macromolecular structure and any change of their network may cause significant structural and functional alterations. Similarly, to the approaches for investigating H–bond network, one can apply simple protocol of the structures of the wild type and mutant or to use snapshot delivered via MD. In both cases, one can use VMD,[70] Chimera,[71] PyMOL,[71] and RASMOL[72] to analyze the H–bond network changes caused by mutations.

2.6. *Electrostatic forces*

Electrostatic forces are frequently omitted from the analysis of effect of mutations, however, recently it was demonstrated that the changes of electrostatic forces due to mutations can be used to discriminate pathogenic from benign mutations.[73] There are no many resources available for modeling electrostatic forces in molecular biology with the prominent exception of DelPhiForce.[74,75] This tool allows for modeling the electrostatic forces acting on individual atom(s), residue(s) or molecular partners. Thus, by comparing the forces within wild type residues and mutant, one can see the differences and attribute them to the effect of these differences on the functionality of the corresponding molecule.

2.7. *Prediction of pH-dependence*

The pKa values of titratable groups in macromolecules determine the pH-dependence of their stability, interactions and enzymatic activity.[76–79] Any changes of the pKa's may affect the wild type properties of the corresponding macromolecule.

However, it should be clarified that these pKa changes should be considered in conjunction with the pH at which the corresponding macromolecule is supposed to be functioning. If the pKa's and their changes are outside this functional pH range, it may be considered that they are not relevant for functionality in vivo. There are several tools for predicting pKa's of titratable groups. The short list includes DelPhiPKa,[80,81] MCCE,[82,83] ProPKA,[84] H++,[85] and CPHMD.[86] Among them, DelPhiPKa is the only one that computes pKa's of proteins, RNAs and DNAs.[87]

3. Molecular Effects and Their Linkage with Diseases

Mutation is the permanent change in the nucleotide sequence of DNA, which occurs during its replication or recombination. The mutation results in substitution, deletion or insertion. Many mutations are repaired before protein synthesis occurs, so they do not have any effects on the function of the corresponding biomolecules. Some of the mutations have positive effect on the corresponding organism and they are called beneficial mutations. However, other mutations may drastically reduce the ability of the organism to survive and such mutations are called harmful mutations. If such a change occurs in a fraction of population but not in a single case, the change is termed single nucleotide polymorphism (SNP).[38] Here, we focus on reviewing computational chemistry tools for modeling effects of mutations that results in a change of the amino acid sequence of the corresponding protein, the nonsynonymous mutations.[59] In this section, we will review the effects caused by nonsynonymous mutations on protein stability, binding, dynamics, H–bonds, salt bridges, electrostatic forces, and pH-dependence and their associations with different diseases.[37] In most cases we will illustrate the linkage between the change of above-mentioned properties and diseases taking examples from our own work, however, it should be clarified that our work represent just a tiny fraction of numerous investigations on this topic.

3.1. *Protein stability changes and their association with diseases*

Usually, the change in folding free energy ($\Delta\Delta G$) is used to determine the stability change of a protein due to a mutation. Depending on the sign of $\Delta\Delta G$, one assesses that the mutation stabilizes or destabilizes the protein. It is expected that any deviation of wild type properties, resulting in over stabilized or destabilized protein, will have negative effect of its function. However, there is no "golden rule" or "golden threshold" to indicate that if the magnitude is large than the cut-off, the mutation will abolish protein function and thus most probably will be pathogenic. Perhaps the best thing to do it to impose another measure like if wild type stability changes by more than a particular percentile, then the functionality will be affected. However, since in vast majority of cases, especially in case of genomic-scale studies, the wild type stability is unknown, one simply uses arbitrary cut-off of 1 or 2 kcal/mol. Here, we provide several examples of mutations affecting protein stability and being

associated with diseases. Thus, it was shown that the mutation M35R in spermine synthase destabilizes the protein and causes Snyder-Robinson syndrome.[88] On the opposite side of spectra is the mutation H101Q in CLIC2 protein, which is shown to increase CLIC2 protein stability and is associated with mental disorder.[89,90] Most frequently mutations destabilize the corresponding protein[91] and cause various diseases.[38] Some of the diseases are Alzheimer's disease,[92] Salt & Pepper syndrome[93] and Snyder-Robinson syndrome.[94] Furthermore, mutations in *LMNA* gene, which are associated with muscular disease,[95] mutations in retinal proteins causing retinal disease[96] and mutations in prion protein associated with prion disease,[97] are also examples of mutations that destabilize the corresponding protein. Similarly, the destabilizations of protein due to mutations are also seen in many neuro-degenerative diseases such as Parkinson's disease.[98] A particular example of destabilizing mutation causing Snyder-Robinson syndrome is shown in Fig. 3(a).

3.2. *Mutation affecting protein-protein and protein-DNA binding and their linkage with disease*

Mutations also have an effect on protein–protein affinity, which can be calculated via the change of protein binding free energy.[58] Many diseases are caused by altered protein–protein, protein–RNA and protein–DNA interactions. It was repeatedly demonstrated that cancer mutations frequently occur at protein binding interfaces.[99] Likewise, mutations in SLC12A3 gene, G439S and G741, which cause Gitelman syndrome are also associated with the change of binding energy.[100] Mutation G56S in spermine synthase destabilize the homo-dimer and is associated with Snyder-Robinson syndrome.[88] Many Rett syndrome-causing mutations are causing alteration of protein-DNA binding.[38,101] Another example are mutations in MLI2 gene, W5065X and R5179H, which leads to Kabuki Syndrome and affect protein binding.[102] A particular example is shown in Fig. 3(b).

3.3. *Mutations affecting protein conformation and linkage with disease*

Proteins function by sampling various conformations. If a mutation affects the conformational space of a protein, it is expected that the function will be affected as well. For example, the mutation L61P in CEP63 protein, which causes aneuploidy and solid tumors in humans,[103] is predicted to increase protein flexibility. Two mutations in the human protoporphyrinogen oxidase (hPPO) gene associated with variegate porphyria, R59Q and R59G, were demonstrated to change hPPO ability to sample relevant conformations.[104] A mutation, L427P, in dystrophin protein was also demonstrated to cause change in conformational flexibility, and to be associated with Becker muscular dystrophy.[105] Mutations in KDM5C gene, N142S and R108W, which are associated with X-linked mental retardation and syndrome Claes–Jensen type disease also alter conformational dynamics of corresponding proteins.[106] A particular example is shown in Fig. 3(c).

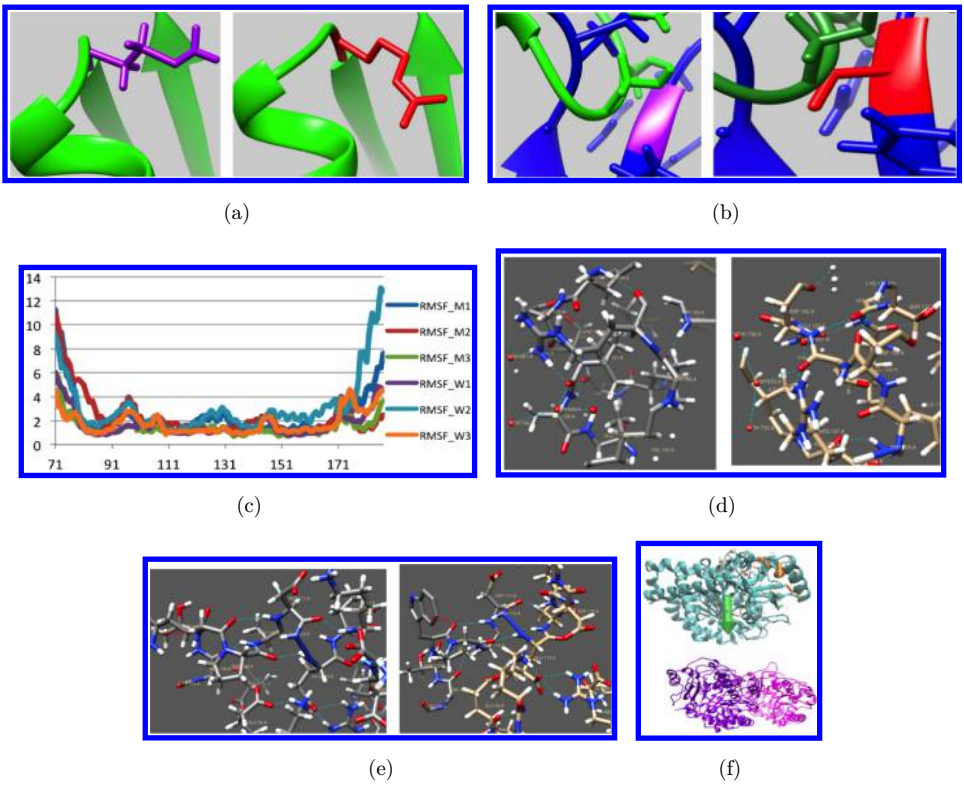

Fig. 3. (Color online) Molecular effects of mutations linked with different diseases: (a) part of spermine synthase zoomed at wild type Met35 (left) and mutant Arg35 (right); mutation is associated with Snyder–Robinson syndrome; (b) part of spermine synthase homo-dimer (chain A in blue and chain B in green) zoomed at wild type Glu56 (left) and mutant Ser56 (right); mutation is associated with Snyder–Robinson syndrome; (c) RSMF of KDM5C ARID domain for mutation R108W based on 100ns MD simulations; mutation is associated with X-linked mental retardation, the syndrome Claes–Jensen type disease; (d) region of wild type I150 (left) and mutant T150 (right) in spermine synthase; the H–bonds are shown in blue dotted lines; one can appreciate the change of H–bond network caused by the mutation; mutation is causing Snyder–Robinson syndrome; (e) Region of wild type R108 (left) and mutant W108 (right) in KDM5C protein demonstrating removal of salt bridge; mutation is associated with syndrome Claes–Jensen type disease; (f) The change of electrostatic forces on kinesin residues bound to tubulin dimer calculated by DelphiForce due to mutation E253K.

3.4. *Mutations affecting H–bond and linkage with diseases*

H–bonds networks are very important for the biological functions of macromolecules. So mutations resulting in removal or addition of hydrogen donor or acceptor have significant effect on structural integrity.[107] It has been investigated that a mutation in CEP63 gene, L61P, resulting in change in H–bond has effect on protein flexibility and associated with aneuploidy and solid tumors.[108] Likewise, the mutation R595W in CPT1 protein results in a change of H–bond network, has effect on protein structure and is associated with mitochondrial fatty acid oxidation disorder.[109] Mutations in multiple sites of amyloid-β peptide causing change of H–bond network

are associated with the Alzhemier's Disease.[110] Another well studied effect of I150T mutation in spermine synthase which causes Snyder Robinson syndrome is predicted to alter the H–bond network in the active site.[111] A particular case is shown in Fig. 3(d).

3.5. *Mutations affecting salt bridges and causing diseases*

Salt bridges are also very important for protein stability and function. Thus, an autistic spectrum disorder is attributed to a salt-bridge deleting mutation, R102Q, in neuronal calcium sensor-1(NCS-1) protein.[108] The syndromic Claes–Jensen-type disease is predicted to be caused by removal of salt-bridge in KDM5C protein due to the mutation R108W.[106] Similarly, the mutation R148H in a globular prion protein associated with Creuzfeldt–Jakob disease affects salt bridge interactions.[111,112] Another mutation, D178N, in the same protein and altering salt bridge interaction, causes fatal familial insomnia (FFI).[110,113] A particular example is shown in Fig. 3(e).

3.6. *Mutations altering electrostatic forces and causing diseases*

Electrostatic forces are very important to understand various molecular phenomena, especially those involving long-range interactions.[114] One can analyze the change of electrostatic forces between wild type and mutant and from there to assess the linkage with diseases. For example, mutation L249Q in kinesin-3 family KIF1A and Y276C in kinesin-1 families showed significant change of electrostatic forces. Both mutations are associated with diseases as Parkinsonism and peripheral neuropathy.[73] A particular example is shown in Fig. 3(f).

3.7. *Mutations affecting pH- dependence and linkage with disease*

pH is an important regulatorily factor for macromolecular function and mutations affecting pH-dependent characteristics of protein may be deleterious. It was demonstrated that H101Q mutation in CLIC2 protein, which is associated with X-linked intellectual disability (ID) affects pH-dependence.[89] Mutations affecting several residues of TTR gene are associated with human amyloidosis and have effect on pH-dependence as well.[115] Mutations in multiple sites of amyloid-β peptide aggregation associated with Alzheimer's disease are also pointed out to alter the pH-dependence.[116]

4. Conclusions

In this work, different approaches of computational chemistry and their applications to model the effects of mutations and their association with diseases were reviewed. The review focuses on tools of computational chemistry; therefore, it is limited to methods that use structural information of any kind. Many widely used approaches to predict pathogenic mutations are not included simply because they are based on

machine learning and utilize evolution information (residue conservation, frequency of mutation, etc.). Thus, the review target computational chemistry audience and attempts to bridge computational chemistry with personalized medicine and precise diagnostics.

It should be outlined that the methods mentioned in this review do not represent exclusive list of all resources available. Here, we attempted to provide only a short list of tools that in our opinion are frequently used to predict effects of mutations of macromolecular stability, binding and conformational dynamics, along with some structural features and pH-dependence. Furthermore, the examples provided are only for illustration and were taken from authors works, while clearly realizing that many other researchers contributed to such kind of investigations as well.

Acknowledgment

The work was supported by Grants from NIH, Grant Numbers R01GM093937 and R01GM125639.

References

1. Grant GH, Richards WG, *Computational Chemistry*, Oxford University Press, Oxford, 1995.
2. Cramer CJ, *Essential of Computational Chemistry*, 2nd ed., Wiley, Chichester, England, 2004.
3. Jensen F, *Introduction to Computational Chemistry*, John Wiley & Sons, Queensland, Australia, 1999.
4. Friesner RA, Ab initio quantum chemistry: Methodology and applications, *Proc Natl Acad Sci* **102**(19):6648–6653, 2005.
5. Thiel W, Semi-empirical quantum-chemical methods, *WIREs Comput Mol Sci* **4**:145–157, 2014.
6. Atkins PW, Friedman RS, *Molecular Quantum Mechanics*, Oxford University Press, Oxford, 1997.
7. Colins MA, Bettens RPA, Energy – based molecular fragmentation methods, *Chem Rev* **115**(12):5607–5642, 2015.
8. Jonshon ER, Becke AD, A post Hartree-Fock model of intermolecular interaction, *J Comput Phys* **123**(2):024101, 1995.
9. Hobza P, Sponer J, Reschel T, Density functional theory and molecular clusters, *J Comput Chem* **16**(11):1315–1325, 1195.
10. Hammond BL, Lester WA, Reynolds PJ, *Monte Carlo Methods in Ab Initio Quantum Chemistry*, World Scientific, Singapore, 1994.
11. Godzik A, Knowledge-based potentials for protein folding: What can we learn from known protein structures, *Structure* **4**(4):363–366, 1996.
12. Hou T *et al.*, Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking, *J Comput Chem* **32**(5):866–877, 2011.
13. Klepeis JL *et al.*, Long-timescale molecular dynamics simulations of protein structure and function, *Curr Opin Struct Biol* **19**(2):120–127, 2009.

14. Piana S, Klepeis JL, Shaw DE, Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations, *Curr Opin Struct Biol* **24**:98–105, 2014.
15. Wang X, Huang SY, Integrating bonded and nonbonded potentials in the knowledge-based scoring function for protein structure prediction, *J Chem Inf Model* **59**(6):3080–3090, 2019.
16. Murphy RB, Philipp DM, Friesner RA, A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments, *J Comput Chem* **21**(6):1442–1471, 2000.
17. Chavent M, Duncan AL, Sansom MS, Molecular dynamics simulations of membrane proteins and their interactions: From nanoscale to mesoscale, *Curr Opin Struct Biol* **40**:8–16, 2016.
18. Hospital A, Goni JR, Molecular dynamics simulations: Advances and applications, *Adv Appl Bioinf Chem* **8**:37–47, 2015.
19. Kappel K, Miao Y, McCammon JA, Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor, *Q Rev Biophys* **48**(4):479–487, 2015.
20. Sneha P, Doss CG, Molecular dynamics: New frontier in personalized medicine, *Adv Protein Chem Struct Biol* **102**:181–224, 2016.
21. Bucher D, Rothlisberger U, Molecular simulations of ion channels: A quantum chemist's perspective, *J Gen Physiol* **135**(6):549–554, 2010.
22. Young DC, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons, Chichester, England, 2001.
23. Doll J, Freeman DL, Monte Carlo methods in chemistry, *IEEE Comput Sci Eng* **1**(1):22–32, 1994.
24. Yamamori Y, Kitao A, MuSTAR MD: Multi-scale sampling using temperature accelerated and replica exchange molecular dynamics, *J Chem Phys* **139**(14):145105, 2013.
25. Peng Y *et al.*, Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver, *Bioinformatics* **44**(5):779–786, 2018.
26. Wang C *et al.*, Recent developments and applications of the MMPBSA method, *Front Mol Biosci* **4**:87, 2018.
27. Ben-Naim A, Statistical potentials extracted from protein structures: Are these meaningful potentials? *J Mol Biol* **257**(2):457–469, 1996.
28. Cossio P *et al.*, A simple and efficient statistical potential for scoring ensemble of protein structure, *Sci Rep* **2**:351, 2012.
29. Sippl MJ, Knowledge-based potentials for proteins, *Curr Opin Struct Biol* **5**(2):229–235, 1995.
30. wwPDB consortium, Protein Data Bank: The single global archive for 3D macromolecular structure data, *Nucleic Acids Res* **47**(D1):D520–D528, 2019.
31. Jankauskaite J *et al.*, SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation, *Bioinformatics* **35**(3):462–469, 2019.
32. Kucukkal TG, Alexov E, Structural, dynamical, and energetical consequences of rett syndrome mutation R133C in MeCP2, *Comput Math Method* **2015**:746157, 2015.
33. Liu Z *et al.*, PDB-wide collection of binding data: Current status of the PDBbind database, *Bioinformatics* **31**(3):405–412, 2015.
34. Zhang N *et al.*, PremPDI estimates and interprets the effects of missense mutations on protein-DNA interactions, *PLoS Comput Biol* **14**(12):e1006615, 2018.

35. Kucukkal TG *et al.*, Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins, *Curr Opin Struct Biol* **32**:18–24, 2015.
36. Stefl S *et al.*, Molecular mechanisms of disease-causing missense mutations, *J Mol Biol* **425**(21):3919–3936, 2013.
37. Zhang Z *et al.*, Analyzing effects of naturally occurring missense mutations, *Comput Math Methods Med* **2012**, 2012.
38. Kucukkal TG *et al.*, Impact of rett syndrome mutations on MeCP2 MBD stability, *Biochemistry* **54**(41):6357–6368, 2015.
39. Rao SN *et al.*, Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin, *Nature* **328**(6130):51–54, 1987.
40. Seeliger D, de Groot BL, Protein thermostability calculations using alchemical free energy simulations, *Biophys J* **98**(10):2309–2316, 2010.
41. Steinbrecher T *et al.*, Free energy perturbation calculations of the thermodynamics of protein side-chain mutations, *J Mol Biol* **429**(7):923–929, 2017.
42. Tidor B, Karplus M, Simulation analysis of the stability mutant R96H of T4 lysozyme, *Biochemistry* **30**(13):3217–3228, 1991.
43. Giollo M *et al.*, A method using residue interaction networks to improve prediction of protein stability upon mutation, *BMC Genomics* **15**:S7, 2014.
44. Getov I, Petukh M, Alexov E, SAAFEC: Predicting the effect of single point mutations on protein folding free energy using a knowledge-modified MM/PBSA approach, *Int J Mol Sci* **17**(4):512, 2016.
45. Gromiha MM *et al.*, ProTherm, thermodynamic database for proteins and mutants: Developments in version 3.0, *Nucleic Acids Res* **30**(1):301–302, 2002.
46. Pires DE, Ascher DB, Blundell TL, Duet: A server for predicting effects of mutations on protein stability using an integrated computational approach, *Nucleic Acids Res* **42**:W314–W319, 2014.
47. Schymkowitz J *et al.*, The foldx web server: An online force field, *Nucleic Acids Res* **33**:W382–W388, 2005.
48. Capriotti E, Fariselli P, Casadio R, Predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res* **33**:W306–W310, 2005.
49. Worth CL, Preissner R, Blundell TL, SDM-A server for predicting effects of mutations on protein stability and malfunction, *Nucleic Acids Res* **39**:W215–W222, 2011.
50. Parthiban V, Gromiha MM, Schomburg D, Cupsat: Prediction of protein stability upon point mutations, *Nucleic Acids Res* **34**:W239–W242, 2006.
51. Masso M, Vaisman II, AUTO-MUTE: Web-based tools for predicting stability changes in proteins due to single amino acid replacements, *Protein Eng Des Sel* **23**(8):683–687, 2010.
52. Gromiha MM *et al.*, ProTherm, version 2.0: Thermodynamic database for proteins and mutants, *Nucleic Acids Res* **28**(1):283–285, 2000.
53. Pezeshgi Modarres H, Mofrad MR, Sanati-Nezhad A, ProtDataTherm: A database for thermostability analysis and engineering of proteins, *PLoS One* **13**(1):e0191222, 2018.
54. Lonquety M *et al.*, SPROUTS: A database for the evaluation of protein stability upon point mutation, *Nucleus Acids Research* **37**(Database issue):D374–D379, 2009.
55. Tajielyato N, Alexov E, Modeling pKas of unfolded proteins to probe structural models of unfolded state, *Journal of Theoretical and Computational Chemistry* **18**(4), 2019.
56. Dehouck Y *et al.*, Beatmusic: Prediction of changes in protein-protein binding affinity on mutations, *Nucleic Acids Res* **41**:W333–W339, 2013.
57. Li M *et al.*, MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions, *Nucleic Acids Res* **44**(W1):W494–W501, 2016.

58. Petukh M, Li M, Alexov E, Predicting binding free energy change caused by point mutations with knowledge-modified MM/PBSA method, *PLoS Comput Biol* **11**: e1004276, 2015.
59. Peng Y *et al.*, Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver, *Bioinformatics* **34**:5, 2018.
60. Wang R *et al.*, The PDBbind database: Methodologies and updates, *J Med Chem* **48**(12):4111–4119, 2005.
61. Xiong P *et al.*, Assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts, *J Mol Biol* **429**(3):426–434, 2017.
62. Pires DEV *et al.*, mCSM: Predicting the effects of mutations in proteins using graph-based signatures, *Bioinformatics* **30**:335–342, 2013.
63. Witvliet DK *et al.*, ELASPIC web-server: Proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity, *Bioinformatics* **32**:1589–1591, 2016.
64. Nelson MT *et al.*, NAMD: A parallel, object-oriented molecular dynamics program, *Int J High Perform Comput Appl* **10**:251–268, 1996.
65. Brooks BR *et al.*, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J Comput Chem* **4**:187–217, 1983.
66. Case DA *et al.*, The Amber biomolecular simulation programs, *J Comput Chem* **26**:1668–1688, 2005.
67. Christen M *et al.*, The GROMOS software for biomolecular simulation: GROMOS05, *J Comput Chem* **26**:1719–1751, 2005.
68. Berendsen HJC, van der Spoel D, van Drunen R, GROMACS: A message-passing parallel molecular dynamics implementation, *Comput Phys Commun* **91**:43–56, 1995.
69. Ponder J, Richards F, TINKER molecular modeling package, *J Comput Chem* **8**:1016–1024, 1987.
70. Humphrey W, Dalke A, Schulten K, VMD: Visual molecular dynamics, *J Mol Graph* **14**(1):33–38, 27–28, 1996.
71. Pettersen EF *et al.*, UCSF chimera — A visualization system for exploratory research and analysis, *J Comput Chem* **25**:1605–1612, 2004.
72. Sayle RA, White EJM, RASMOL: Biomolecular graphics for all, *Trends Biochem Sci* **20**(9):374–376, 1995.
73. Li L *et al.*, Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations, *Sci Rep* **7**(1):8237, 2017.
74. Li L, Chakravorty A, Alexov E, DelPhiForce, a tool for electrostatic force calculations: Applications to macromolecular binding, *J Comput Chem* **38**(9):584–593, 2017.
75. Li L *et al.*, DelPhiForce web server: Electrostatic forces and energy calculations and visualization, *Bioinformatics* **33**(22):3661–3663, 2017.
76. Alexov E, Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure, *Eur J Biochem* **271**(1):173–185, 2004.
77. Mitra RC, Zhang Z, Alexov E, In silico modeling of pH-optimum of protein-protein binding, *Proteins* **79**(3):925–936, 2011.
78. Onufriev AV, Alexov E, Protonation and pK changes in protein-ligand binding, *Q Rev Biophys* **46**(2):181–189, 2013.
79. Talley K, Alexov E, On the pH-optimum of activity and stability of proteins, *Proteins* **78**(12):2699–2706, 2010.

80. Pahari S *et al.*, DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate, *Proteins* **86**(12):1277–1283, 2018.
81. Wang L, Zhang M, Alexov E, DelPhiPKa web server: Predicting pKa of proteins, RNAs and DNAs, *Bioinformatics* **32**(4):614–615, 2016.
82. Gunner MR, Zhu X, Klein MC, MCCE analysis of the pKas of introduced buried acids and bases in staphylococcal nuclease, *Proteins* **79**(12):3306–3319, 2011.
83. Zhang J, Gunner MR, Multiconformation continuum electrostatics analysis of the effects of a buried Asp introduced near heme a in Rhodobacter sphaeroides cytochrome c oxidase, *Biochemistry* **49**(37):8043–8052, 2010.
84. Rostkowski M *et al.*, Graphical analysis of pH-dependent properties of proteins predicted using PROPKA, *BMC Struct Biol* **11**:6, 2011.
85. Anandakrishnan R, Aguilar B, Onufriev AV, H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, *Nucleic Acids Res* **40**(Web Server issue):W537–W541, 2012.
86. Wallace JA, Shen JK, Predicting pKa values with continuous constant pH molecular dynamics, *Methods Enzymol* **466**:455–475, 2009.
87. Wang L, Li L, Alexov E, pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa, *Proteins.* **83**(12):2186–2197.
88. Peng Y *et al.*, Revealing the effects of missense mutations causing snyder-robinson syndrome on the stability and dimerization of spermine synthase, *Int J Mol Sci* **17**(1):77, 2016.
89. Takano K *et al.*, An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity, *Hum Mol Genet* **21**(20):4497–4507, 2012.
90. Witham S *et al.*, A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics, *Proteins* **79**(8):2444–2454, 2011.
91. Shi Z, Moult J, Structural and functional impact of cancer-related missense somatic mutations, *J Mol Biol* **413**:495–512, 2011.
92. Dolzhanskaya N *et al.*, A novel p.Leu(381)Phe mutation in presenilin 1 is associated with very early onset and unusually fast progressing dementia as well as lysosomal inclusions typically seen in Kufs disease, *J Alzheimers Dis* **39**:23–27, 2014.
93. Boccuto L *et al.*, A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation, *Hum Mol Genet* **23**:418–433, 2014.
94. Zhang Z *et al.*, Computational analysis of missense mutations causing Snyder-Robinson syndrome, *Hum Mutat* **31**:1043–1049, 2010.
95. Zwerger M *et al.*, Myopathic lamin mutations impair nuclear stability in cells and tissue and disrupt nucleo-cytoskeletal coupling, *Hum Mol Genet* **22**:2335–2349, 2013.
96. Rakoczy EP *et al.*, Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations, *J Mol Biol* **405**:584–606, 2011.
97. Prusiner S, Molecular biology of prion disease, *Science* **252**:1515–1522, 1991.
98. Nuytemans K *et al.*, Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: A mutation update, *Hum Mutat* **21**:763–780, 2010.
99. Nishi H *et al.*, Cancer missense mutations alter binding properties of proteins and their interaction networks, *PLoS Comput Biol* **8**(6):e66273, 2013.
100. De Jong JC *et al.*, Functional expression of mutations in the human NaCl cotransporter: Evidence for impaired routing mechanisms in Gitelman's syndrome, *J Am Soc Nephrol* **13**:1442–1448, 2002.

101. Yang Y *et al.*, Binding analysis of methyl-CpG binding domain of MeCP2 and Rett syndrome mutation, *ACS Chem Biol* **11**(10):2706–2715, 2016.
102. Ng SB *et al.*, Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome, *Nat Genet* **42**:790–793, 2010.
103. Kumar A, Purohit R, Computational investigation of pathogenic nsSNPs in CEP63 protein, *Gene* **503**:75–82, 2012.
104. Wang B *et al.*, Quantitative structural insight into human variegate porphyria disease, *J Biol Chem* **17**:11731–11740, 2013.
105. Acsadi G *et al.*, Novel mutation in spectrin-like repeat 1 of dystrophin central domain causes protein misfolding and mild Becker muscular dystrophy, *J Biol Chem* **287**:18153–18162, 2012.
106. Peng Y *et al.*, Mutations in the KDM5C ARID domain and their plausible association with syndromic Claes-Jensen-Type disease, *Int J Mol Sci* **16**(11):27270–27287, 2015.
107. Horowitz S, Trievel RC, Carbon-oxygen hydrogen bonding in biological structure and function, *J Biol Chem* **287**:41576–41582, 2012.
108. Zhu Y *et al.*, R102Q mutation shifts the salt-bridge network and reduces the structure flexibility of human neuronal calcium sensor protein, *J Phys Chem* **118**(46):13112, 2014.
109. Fontaine M *et al.*, A novel mutation in CPT1A resulting in Hepatic CPT deficiency, *JIMD Rep* **6**:7–14, 2012.
110. Attanasio F *et al.*, Carnosine inhibits Abeta42 aggregation by perturbing the H-Bond network in and around the central hydrophobic cluster, *Chembiochem* **14**:583–592, 2013.
111. Kamp MW, Daggett V, The consequences of pathogenic mutations to the human prion protein, *Protein Eng Des Sel* **22**(8):461–468, 2009.
112. Pastore M *et al.*, Creutzfeldt-Jakob disease (CJD) with a mutation at codon 148 of prion protein gene: Relationship with sporadic CJD, *Am J Pathol* **167**(6):1729–1738, 2005.
113. Petersen RB *et al.*, Effect of the D178N mutation and the Codon 129 polymorphism on the metabolism of the prion protein, *J Biol Chem* **271**(21):12661–12668, 1996.
114. Li C *et al.*, DelPhi suite: New developments and review of functionalities, *J Comput Chem* **40**(28):2502–2508, 2019.
115. Yokoyama T *et al.*, Hydrogen-bond network and pH sensitivity in transthyretin: Neutron crystal structure of human transthyretin, *J Struct Biol* **177**:283–290, 2012.
116. Garai K, Baban B, Frieden C, Self-association and stability of the ApoE isoforms at low pH: Implications for ApoE-lipid interactions, *Biochemistry* **50**:6356–6364, 2011.