Fast electrostatic calculations in energy analysis : method development, applications in investigating organic and organometalic crystals, and to protein/ligand complexes

Vedran Vuković

Collegium Sciences et Technologies

Ecole Doctorale C2MP

THÈSE

Présentée pour l'obtention du titre de

Docteur de l'Université de Lorraine

en Physique

par

Vedran Vuković

Calculs électrostatiques rapides dans l'analyse énergétique

Développement de la méthode, applications aux études de cristaux organiques et organométalliques et de complexes protéine/ligand

Soutenance publique le 20 septembre 2021 à 14h00, Salle des Actes
Faculté des Sciences et Technologies

Membres du jury :

Rapporteurs, madame et monsieur :

Paulina Dominiak	Professeur, Faculté de Chimie, Université de Varsovie
Emmanuel Fromager	Professeur, LCQ, Université de Strasbourg

Examinateurs, mesdames et messieurs :

Paulina Dominiak	Professeur, Faculté de Chimie, Université de Varsovie, Pologne
Bertrand Fournier	Maître de Conférences, IGPS, Université Paris-Saclay
Emmanuel Fromager	Professeur, LCQ, Université de Strasbourg
Alessandro Genoni	CNRS Chargé de Recherche, LPCT, Université de Lorraine
Francesca Ingrosso	Maître de Conférences hC, LPCT, Université de Lorraine
Christian Jelsch*	CNRS Directeur de Recherche, CRM², Université de Lorraine

Invité, monsieur :

Claude Didierjean	Docteur, CRM², Université de Lorraine
Benoît Guillot	Professeur, CRM², Université de Lorraine

* directeur de thèse

Remerciements

…fregit…

Je prie le lecteur de ne jamais dire qu'il a lu *ma* thèse. Bien que c'était moi qui ai passé de longues nuits à veiller à ce que chaque mot trouve sa place – et j'ai failli à bien faire même cela – la réussite de ma thèse dépendit de circonstances hors de mon contrôle. Il n'est pas juste de m'attribuer la gloire non méritée, et il sera encore moins juste d'oublier ceux qui ont contribué à cette œuvre. Ceci dit, il faudra un livre aussi grand que cette thèse pour rendre le vrai tribut, et un homme moins bête et oublieux pourrait l'achever. Mes excuses les plus profondes à ceux qui j'oubliai.

La thèse est inséparable de son directeur, de l'homme qui a conçu son but et son sujet. Merci, Christian, de m'avoir guidé par les chemins durs de la recherche sur les collines électrostatiques, de m'avoir conseillé aux moments quand le chemin semblait disparaître devant moi, d'avoir soutenu mes propos à la fois fous et impertinents, mon tempérament à la fois sévère et têtu.

Merci Benoît d'avoir passé un temps énorme à discuter avec moi. Ce n'était jamais du temps perdu, et j'étais toujours calmé de savoir que je peux compter sur toi. J'espère de ne pas avoir abusé de ta bonté, qui te poussait parfois à négliger tes propres obligations urgentes pour m'aider.

Merci Claude pour nos longues discussions, à fois sur la matière scientifique, à fois sur tout le reste qu'une vie apporte. J'apprécie l'esprit chaleureux et ouvert de nos échanges, qui m'a permis de considérer l'intégralité de mes circonstances, et pas juste le monde restreint et quotidien du laboratoire.

Merci à mes collègues permanents du laboratoire : à Dominik pour son soutien administratif et aupres de l'école doctorale, sa patience à répondre à toutes mes nombreuses questions et ses conseils scientifiques sur les propositions de projets ; à Maxime parce qu'il a partagé ses avis sur les sujets scientifiques et politiques ; à Emmanuel qui m'a appris comment mener les expériences de la diffraction et qui avait toujours du temps à m'écouter ; à Frédérique pour son regard positif et chaleureux et un conseil fructueux ; à Anne, Christine et Valérie pour leur aide avec les démarches administratives et leur patience, surtout quand je ne parlais pas français ; à Abdelatif, Arnaud, Axel, Bruno, El-Eulmi, Élodie, Emmanuel, Enrique, Massimo, Nicolas, Pierrick, Remi, Sandrine, Sébastien et Yohann pour chaque rencontre et échange ; à tous pour un accueil exceptionnel et un soutien incomparable.

A great many thanks to the PhD and post-doc fellows from my lab : à Théo qui était mon premier soutien quand je ne pouvais pas dire un mot de français, j'ai bien aimé nos discussions sur la science et la programmation, sur les projets en commun, sur la France et l'avenir ; à Julien pour tout son aide les premiers jours de ma thèse, qui m'a permis de mieux me repérer en France et naviguer dans les nombreux services publics ; to വിഷ്ണു, who has grown from the confused Indian boy alone in a big new country to a confident man he is now, and who has allowed me to witness this great transformation, even if from afar ; à Eva que je connais injustement peu, et qui est autant brillante qu'ouverte ; à Ridvan pour le travail que nous avons fait ensemble pour le séminaire de l'école doctorale ; à Aki-

hiro, Amatassalâm, Amidou, Amira, Hasan, Krzysztof, Ibrahima, Matthieu, Morgane, Rahul, Samah, Serigne, Simon, et à tous ceux que j'aurais pu honteusement oublier.

Ich bin unendlich dankbar für die Zeit, die ich in Regensburg verbracht habe, die beste meiner gesamnten Doktorarbeit. Vielen Dank, Michael, für deine Freundschaft und Unterstützung. Es ist ein Vergnügen, mit dir zu arbeiten. Danke auch an meine Bürokolleginnen, Birgitt, Charmagne und Sabine, mit denen jeder Tag Spaß gemacht hat. Danke auch an Christoph und Lena für die tolle Zeit in Wien und Regensburg. Herzlichen Dank an Monika, die die beste Vermieterin in der ganze Welt ist (keine Übertreibung). Ich muss auch Ari bedanken, der meine Angst vor Hunden geheilt hat. Liebe Danke an all die anderen Leute, die mein Aufenhalt verschönert haben: Matilde, Anna, Coralie, Gábor, Ivanu, Lari, Lorenzo e sua moglie Kristin(a), Martin, der andere Michael, der dritte Michael, Nikolaus, Phileas, Stephanie, Verena, und weitere Menschen, deren Namen ich nicht mehr erinnern kann oder nie gelernt habe – insbesondere eine Person aus der letzte Kategorie.

Ich darf auch Christian und Christopher nicht vergessen: wir hatten in Wien viele schöne lange Abends zusammen verbracht und das Naturwissenschaftlichen Museum besucht (da wären noch Michael, Lena und Christoph). Danke auch an Theo und Beate für ihre Gastfreundschaft in Saarbrücken, und die vielen folgenden Gespräche.

Many, many thanks to my colleagues outside of my lab, whom I met on various occasions. First and foremost is Anna, who was the sunlight during the earliest days of my PhD. Après vient Sajatan, qui m'a lié d'amitié aisément, et qui était toujours là quand je passais à Strasbourg, et je ne peux pas oublier quand il m'a accompagné à 5h du matin au bus annulé. Thanks to so many others: Alessandro, Bertrand, Delia, Emmanuel, Francesca, Paulina, Radek, Roberto, Vincent, and also Alekos, Ali, Alix, Antonio, Arthur, Aurianne, Bjarke, Cécile, Cécilia, Claude, Clément, Dalia, Дарья, David, Erna, Florian, Florence, Gautier, Gaukar, Илона, Jiaqi, Людми́ла, Malte, Marc, Мария, Мария, Martin, Mateusz, Максим, Natalia, Paolo, Rabah, Серге́й, شعيب, Taïcія, Valentin, Valentin, Victor, Владимир, Zarina, and many others.

Merci Christine pour vos sourires et votre joie que vous avez partagée à chaque réunion. Merci George et Juliette pour le soutien administratif, pour tout ce que vous avez fait pour faciliter le doctorat.

Hvala Kreši i Tomi na suradnji i razgovoru, a prvenstveno na prijateljstvu. Hvala Vladislavu za prijateljske razgovore i dobrodošlicu, kao i za mogućnost držanja vježbi na Kemometrici. Hvala svim mojim profesorima i kolegama studentima koje sam viđao kad sam bio u Hrvatskoj: Andreji, Biserki, Damjanu, Darku, Dominiku, Dubravki, Dunji, Ediju, Ernestu, Franu, Gloriji, Gregoru, Grguru, Jerku, Josipu, Ivani, Ivanu, Ivanu, Ivici, Katarini, Karlu, Kristini, Marijani, Marini, Mateju, Matiji, Mihaelu, Nataši, Nei, Nikoli, Petru, Siniši, Stjepanu, Tinu, Vinku, Vladimiru, Zoe, Željki, Željku, bez posebne namjere da ih izdvojim. Hvala i Josipu na našem dugo planiranom susretu u Francuskoj.

J'étais béni des amitiés qui ont marqué mon séjour en France, surtout ces derniers mois. Je dois remercier, parmi d'autres, Agathe, Alexandre, Alexandre, Arthur, Antoine, Anthony, Augustin, Aurélie, Benoît, Benoît, Claire, Charles, Christophe, Clémence, David, Daniel, Daniël, Domitille, Edouard, Edwige, Eloi, Eloi, Gaétan, Grégoire, Grégoire, Fabrice, François, François-Joseph, François-Xavier, Hélène, Jean, Jean-Baptiste, Jean-Gabriël, Jean-Marie et sa femme, Laurent, Marc, Marie, Marie-Claire, Martin, Mélanie, Nicolas, Laurie, Luc-Henri, Maxence, Olivier, Pascale, Paul-Emmanuel, Paul-Emmanuel, Philippe, Pia, Pierre, Pierre et sa famille, Romary, Sébastien, Simon, Sophie, Thierry, Théodor, Victor, Vincent, Vincent, Vincent, Yv, Alexis et Sylvain, dont je crains avoir abusé la bonté, Kévin avec la tête dure et le grand cœur, Johann qui est toujours prèt a sourire, Martin qui trouve de temps pour

moi malgré ses nombreuses occupations. Merci particulièrement à Charles, Charles, Clément, Louis et surtout Arnoud. Je suis vraiment désolé car je n'ai pas retenu les prénoms de tous.

Merci aussi à Audrey, Andrzej, Benoît, Matthieu, Yennifer, Yusuf et plein d'autres ! Merci à Kevin, Kevin, Kevin, Kostia, Matthieu, Stéphane, Sacha, Xavier, et d'autres dont le nom j'ignore.

Un grand merci à Stanislas, Jean-Vianney, Florent et Dominique, desquels j'ai appris un nombre énorme de choses. En particulier, merci à Stanislas pour les nombreux repas partagés où nous discutions d'amour ; à Jean-Vianney, qui a partagé les beautés naturelles, culturelles et historiques de la France, ainsi que plein de conseils pratiques ; à Florent, qui m'a appris le service, qui a patiemment répondu à toutes mes questions (dont la source semble toujours inépuisable), qui a toléré mes impertinences et mes inaptitudes, et qui m'a confié l'apprentissage des autres ; à Dominique pour son accompagnement dans les mystères de la vie, aussi innocents semblent-ils. Quel malheur, de devoir vous abandonner !

Napustiti se čini neizbježno, skoro očekivano. Ponekad se vratiti, posjetiti dobre stare prijatelje, samo da nas mač sudbine opet razdvoji i iznova natjera da idem na daleke staze. Tužan molim ispriku za odsutnost i u isti mah zahvaljujem Alenu, Carlu, Draženu, Dominiku, Dorianu, Frani, Ines, Ivici, Lovri, Mariji, Marijani, Marku, Marti, Mateju, Mihovilu, Mirjam, Mirjani, Miroslavu, Mislavu, Petru, Petru, Petru, Steli, Tinu, Toniju; Kreši za dugačke razgovore o polititci, znanosti i statistici, Marku za svesrdnu pomoć u jednoj od najtežih stvari moga života, Marinu za povjerenje kroz teške trenutke tvoje borbe, Tadiji (ovaj put te nisam zaboravio!), Tomislavu za sve naše dugačke i filozofske rasprave (a isprike za sve krive trenutke).

Posebno hvala društvancu koje sam tu i tamo okupio za kemičarsko/biološke razgovore. Marija, redovito si bila šutljiva, ali nas je tvoj osmijeh uvijek radovao, tvoja je pozitiva svijetlila nad svim problemima koje smo iznosili. Kristin, kad god se ukaže prilika da se ponovno sastanemo, uvijek nastavimo naše prijateljstvo kao da se nikad nismo ni rastali. Nina, inspirirala si me i pomogla mi uvidjeti da mogu i trebam više nego što sam se bojao priznati. Vigore, ne znam kako se svaki put ponovno sastanemo, našim mnogim i velikim razlikama unatoč, ali drago mi je da ih nikad nismo vidjeli kao prepreke.

Velika hvala obiteljskim prijateljima i susjedima, Bosi, Ivi, Ivki, Kseniji, Magali, Nevenu, Dragici za gostoprimstvo i svaku učinjenu uslugu, Jadranki i njenim curama Gabrieli, Zdravki i Rafaeli za prijateljstvo i velikodušnost, za gostoprimstvo kako u Parizu, tako i u Ulmu, Jadranki za njene korisne medicinske savjete, Ivi i Neiri za nenadoknadiva dobročinstva i redovitu spremnost za pomoć, Mariji, Peki i Petri za brižnost, Mariju i Ani za pripomoć i prijenos obrane, Mili za ljubav prema kemiji i za njeno veliko srce, opterećeno tolikim bolima ovoga svijeta, Slavenki za prijateljstvo unatoč problemima, Slobodnoj za sve ispečene kolače i sva dobročinstva, i nemojte da se tko naljuti što nije spomenut, već okrivite moje slabo pamćenje!

Hvala za sva dobročinstva Filku, Miri i Nevenki. Počivali u miru.

اسا. I enjoy your joyful spirit, always willing to throw a bone, a joke. I thanked so many for their warmth and hospitality, but to you I thank for your unending cheer. I hope that life never beats you down and that you bring as much happiness to the world as you have to our lab in these short years that we spent there together.

Артём. You have always wanted to be my friend, even when I behaved stupidly. You taught me that differences can coexist and that the present moment might be as important as an unknown future. If not for you, I would only know some people superficially.

احتشام. Your friendship was invaluable as it strenghtened some of my convictions. I can never forget the nights out in the bar and the long discussions about politics, history, religion and beauty. We have

proven that you leaving Nancy does not impede our friendship: let it remain so even when I leave France.

विशाल. Fifteen books would not suffice to describe the peculiar character that you are. A man who could not like you would have no heart. You were always the last, but you never gave up nor gave in, and you would always manage – no matter what the subject was. Getting to know you is such an honour that I still do not understand what made me worthy of it. Thank you.

Mihai. I always admired your youthly idealism, the kind I just seemed to have lost not long before encountering you. Our conversations sometimes felt like talking to my younger self, and my frustrations were directed – or so my pride likes to wisper – more to my own blindness than to the beam in your eye. I enjoyed thoroughly the emotional and intellectual challenge that our debates were. I am also thankful for your kind spirit and the time spent together for its own sake.

Nicolas. Ta confiance en moi fut un grand honneur, que j'espère ne jamais trahir. Merci pour ta perspective unique et tout le temps partagé, aussi peu était-il !

Paul. Ta simplicité et ta volonté à aller au delà de tes forces, malgré tout, m'ont toujours fortement impressionné. Tu étais un guide fidèle sur la terre touloise, et tu m'as suggéré un lieu loin à visiter qui restera toujours dans mon cœur. Les mots défaillants à dire les secrets de l'âme, il ne me reste que te remercier, pour tout.

Simon. Je ne sais pas pourquoi j'ai passé toutes ces heures te suivant dans les rues de Nancy, t'accompagnant à donner le bonheur aux misérables, mais je sais que je n'aurais pas échangé une seule seconde même pour tout le trésor du monde. Je ne sais pas pourquoi nous sommes amis – qu'est-ce que se cache au fond des cœurs des hommes ? – mais je sais que nous le sommes. Quel don sans prix, pourtant donné gratuitement ! Merci, merci, merci.

Flo. Warum hast du mich befreundet? Könnte man keine Ruhe haben, ohne ewige Bindungen? Wäre es nicht einfacher, wenn ich Regensburg verlassen könnte, und nicht mehr zurückkommen würde? Alles einfach vergessen. Nein. Nach dem Winter kommt der Frühling, und das Feuer einer wahren Freundschaft mag schwinden, aber es wird nie erlöschen. Dafür bin ich dich zu Dank verpflichtet.

Rok. Kako je lako kleti sudbinu – Providnost – pa kakve tek darove ona nosi. Naš slučajan susret otkrio mi je i ukazao na toliko toga što je pokraj mene prolazilo, a da nisam zamjećivao ili nisam htio zamijetiti. Nadam se da je moj udio manji, da će većina blagodati pripasti tebi.

Daniel. Rijetko mi se dogodi da nađem istomišljenike. Naši su se putevi zato ukrstili, ali bi se i razdvojili da nismo bili voljni hoditi sličnim stazama, makar tisuću kilometara daleko. Došlo vrijeme da hodimo bliže!

Kristina. Kolika je blagodat imati jednu tako staloženu i prizemljenu osobu nedaleko! Hvala ti za sve duge razgovore i za slušanje svih mojih beskonačnih priča i problema, za sve tople riječi i bodrenje.

Zrinka. Došli smo do istog mjesta različitim putevima. Nevjerojatno je gledati različitost naših pristupa, pogotovo u filozofksim raspravama, a sličnost rezultata. Kako je slučaj s drugima, tako i s tobom: nisu rasprave prve, već tek najviši kamen na zgradi ljudskosti i prijateljstva. Hvala ti što si je gradila sa mnom.

Antonio. Okolnosti su se promijenile, ali ti si ostao uvijek isti. Staložen, spreman na svaku kulinarsku avanturu, voljan pružiti dobrodošlicu i gostoprimstvo kome god zatreba, kad god zatreba. Hvala ti na svemu što si podijelio, kako na ukusnoj hrani, tako i na toplim riječima. Hvala ti za povjerenje.

Hvala mojoj obitelji i rodbini, koji su uvijek bili kraj mene kad je zagustilo, i koji su s veseljem dočekali svaku moju novost sa studija i iz daleke Francuske: Ivani, Ivanu, Jadranki, Jošku, Maji, Mariji, Marini, Mislavu, Slavki, Tomi, Danielu za uvijek svježu i drukčiju perspektivu, te za ponos kad se ističe mojim rođakom, Ivici, Annie, Natachi i Julie – pour votre soutien et parce que je pouvais toujours

compter sur vous, si loin de la terre natale, Matku za druženje i pomoć u nabavi auta (hvala što si me nagovorio!), Nikši za sve naše kompleksne razgovore, zato što je uvijek inzistirao na tome koliko malo znamo, unatoč mojim uobraženim tvrdnjama u suprotno.

Hvala ti, Marušo, što si me sjetila koliko sam daleko od obitelji. Žao mi je što me udaljenost koja nas razdvaja spriječila da te vidim kako odrastaš, tebe koja si rođena dan nakon početka mog doktorata, da sudjelujem u malim radostima koje male ljude čine bitnijim od velikih. Nadam se da ćeš, prohtije li ti se ikada poći mojim stopama, mudro odvagati i odlučiti. Više od toga, nadam se da ćeš se uvijek radovati, kroz sve dugačke godine koje stoje pred tobom, kao što se smiješiš od prvog dana kad sam te upoznao.

Velika hvala mojim tetama Koviljki i Ljiljani i mom stricu Svetislavu, što su bili potpora i oslonac od malih nogu, što su se uvijek svesrdno radovali mojim uspjesima, što su uvijek bili pri ruci kad bi zatrebalo, pri telefonu kad bih dijelio svoje brige i nedaće, što su me uvijek strpljivo podnosili. Što bih ja bez vas?

Vlatane, uvijek si bio praktičniji od mene, sposoban razumjeti tehničke stvari i upoznat s toliko toga što je uvijek bilo okraj mog obzora. Radeći uz tebe sam se osjećao glupasto radi toga što sam potrošio godine u apstraktnim idejama i naukama i tako se udaljio od konkretnog. Tebi to nije smetalo, ti si samo strogim pogledom i mirnim tonom objasnio što nisam znao, što nisam vidio. Dao si mi da iskusim ljepotu ručnog rada, a uklonio sve teškoće razmišljanja o konstrukciji i izradi. Savjetovao si me u mnogočemu, od kućnih popravaka do auta, o zdravlju i radu. Bio si mi pri ruci kad je zatrebalo. Došao si u Francusku ne zbog Francuske niti zbog puta, nego da podijeliš sa mnom dan obrane, dan koji bi trebao biti najveći i najbitniji u mom životu. Kako pronaći načina da ti se odužim?

Nardi, bio si mi i briga i radost. Nikad ne bih poželio što te spopalo, pa ipak osjećam da sam pridonio, makar propustom. Svjestan, pokušah napraviti što sam mogao i vidio. Drago mi je što si smogao snage i hvala ti za povjerenje da mi se obratiš. Makar vidjeli drukčije, ipak smo uspjeli naći zajednički jezik. Uspio si se oporaviti. Ma dao bih sve doktorate i diplomske, sve studije i nauke, da nikad više ne moram napisati jednu takvu rečenicu! Podigao si se, uzvisio iznad svega što si ikad mislio da možeš. Raduje me tvoja promjena, veseli napredak. Drago mi je vidjeti što postaješ, u što odrastaš. Došao si podijeliti moju obranu sa mnom, volio bih doći i podijeliti tvoju s tobom. Nikad nemoj pomisliti da sam daleko ili visoko: za tebe sam uvijek tu, kakve god udaljenosti nas dijelile. Ostani što jesi, i postani bolji! Za sve, zlo kao i dobro, tebi hvala.

Što reći, sada i ovdje? Kako sažeti ovo sve? Volim vas.

Na kraju svega shvaćam što mi je rečeno kad sam odlazio iz Hrvatske: neki će otići, drugi doći, neki će postati bitniji, drugi sporedni. Uvijek će biti ljudi, a obitelji se uvijek mogu vratiti, ma kako daleko otišao: možda sam vas napustio, ali vi mene niste nikada.

I tako, dragi čitatelju, sada vidiš koliko ovog rada pripada meni: koliko ostane kad se razdijeli na tisuće dijelova, i u svaki utka po jedna duša koja je kroz ovo vrijeme moju dušu dotaknula, i u tom se blagom trenutku u ovo djelo uplela. O nepravde, da ne mogu više od ovoga, kao da ne zaslužujete više! Tišina neka prenese što riječi ne mogu.

Parce mihi, Domine, quia Dalmata sum!

Milenæ et Drazislavo:
honor paternitatis manet in sæcula.

P

Contents

1 General Introduction

1.1 Résumé

Le but principal de cette thèse est d'élaborer une façon de calculer l'énergie d'interaction électrostatique entre les molécules. L'approche la plus simple est de calculer l'énergie entre les deux charges ponctuelles q_A et q_B séparées d'une distance d selon la formule de Coulomb : $q_A q_B / d$.[1]

Cette approche ne prend pas en compte l'énergie dite de pénétration : l'énergie qui provient du chevauchement des densités électroniques causé par la proximité des molécules. Pour résoudre ce problème, il faut d'abord avoir un modèle plus détaillé de la densité électronique atomique dans les molécules. Le modèle multipolaire de Hansen et Coppens est utilisé dans le cadre de cette thèse.[2] Ce modèle ne voit pas les atomes comme étant sphériques. Il inclut un terme de la densité multipolaire qui sert à mieux décrire la densité moléculaire, surtout dans les liaisons covalentes et les paires libres, qu'un modèle à atomes sphériques ne prend largement pas en compte.

Un modèle multipolaire décrit l'ensemble de la densité moléculaire dans la mesure du possible, en prenant en compte les limitations expérimentales. Le calcul d'énergie électrostatique à partir de ce modèle devient plus complexe (équation 1.2) et sera abordé dans le chapitre suivant. La complexité et la répétition du calcul demandent une implémentation informatique, qui sera discutée dans le chapitre 3. Enfin, trois chapitres sont dédiés aux applications possibles de ce logiciel : le premier sur un complexe protéine-ligand, le deuxième sur un contact singulier dans un cristal métallo-organique, et le troisième sur l'énergie électrostatique de la maille cristalline, sur deux exemples différents.

Avant de partir dans ces directions, l'introduction générale aborde aussi la question des autres méthodes similaires. Du côté fondamental, la théorie de la fonctionelle de la densité (la DFT) permet de calculer l'énergie totale (dont l'intégrale 1.2 fait partie) à partir d'une densité calculée pour un système chimique quelconque. L'inconvenient principal de la DFT est l'inaccessibilité des systèmes larges.

Une alternative basée partiellement sur la DFT est la méthode d'énergie des noyaux (ang. *kernel energy method*, KEM). Cette méthode obtient des énergies assez précises pour les larges systèmes, tels que les biomacromolécules. La méthode dévelopée dans le cadre de cette thèse peut également s'appliquer sur les biomacromolécules avec simplicité, n'ayant besoin que d'un transfert de paramètres multipolaires à partir d'une banque de données.

Cette dernière idée peut rappeler un champ de force, qui contient deux parties : les paramètres applicables aux molécules et la méthode de calcul d'énergie. L'analogie est adéquate, sauf que les

[1] Coulomb, 1785, p. 529.
[2] Hansen et Coppens, 1978.

champs de force traditionnels utilisent les charges ponctuelles, ajustées pour bien reproduire des valeurs cibles. Il reste encore du travail pour aboutir à un équivalent de champ de force basé sur une densité de charge multipolaire de Hansen et Coppens.

1.2 *Ubi sum?*

Finding a proper location for this thesis within the myriad specialised fields of modern science seems nigh impossible. Its subject dabbles into mathematics, meanders into physics and programming, makes liberal use of crystallography – a field already at an intersection of so many others – and visits chemistry and biology when necessary. This thesis cannot be placed into any one field easily, so a discussion of where it belongs to is worthwhile.

The fundamental drive behind the thesis project was to calculate electrostatic interaction energy between two molecules in a rapid fashion. Electrostatic energy lies at the heart of the electromagnetic theory. It describes how strongly two charges that do not fluctuate with the passage of time attract or repel each other. If the two charges q_A and q_B are point charges at a distance d, the electrostatic energy between them boils down to a well-known formula derived by Coulomb:[3]

$$E = \frac{q_A q_B}{d} \tag{1.1}$$

Atoms and molecules consist of charged nuclei and electrons. While representing atomic nuclei as point charges is good enough for most practical applications, electrons are more mobile and call for an elaborate depiction. One describes the bulk of a material with a mass density without recourse to finer details such as exact atomic positions. The same logic applies for electrons: one uses *electron density* to show an average of locations where electrons dwell over time.

The literature reports on various basic models to represent atomic and molecular electron densities.[4] Hansen and Coppens developed a sophisticated *multipolar* model to describe the blur of electrons around atomic nuclei based on experimental measurements.[5] Precise electrostatic energies between molecules must rest on a firm basis like the Hansen-Coppens model.

Before Hansen and Coppens, Buckingham introduced a simpler approximation of *multipolar moments* (MM) to calculate electrostatic energy between atoms. His approximation works rather well

[3] Coulomb, 1785, p. 529.

[4] Much of the work on atomic and molecular electron densities relies on the notion of molecular orbitals as linear combinations of atomic orbitals. Thus, from a perspective of a quantum or density functional theory calculation, an electron density is but the square of the wave function that comes out at the end of a calculation. Such results are building blocks for empirical models that describe experimental data, typically from X-ray diffraction studies. The book by Coppens, 1997, pp. 49–71, introduces several molecular electron density models. The most common in structure determination is the independent atom model (Cochran, 1956; Tomiie, 1958) in which the molecules consist of neutral, spherical atoms. A step away from the first assumptions is the spherical κ formalism (Coppens *et al.*, 1979), in which atoms are no longer neutral, and thus no longer independent – the sum of all charges in the crystal must remain zero at all times. The multipolar atom model (Hansen and Coppens, 1978) removes the second assumption (sphericity) from atoms by introducing spherical harmonics to describe non-spherical density. An alternative is Hirshfeld's model (Hirshfeld, 1971) based on similar mathematical principles.

[5] The original paper is Hansen and Coppens, 1978. Section 3.2.4 from Coppens, 1997, pp. 59–67, gives a detailed description of the model and its various facets.

when those atoms are at some distance apart.[6] Atoms interact weakly at such distances since their electron densities do not overlap significantly. One can derive Buckingham's multipolar moments directly from Hansen-Coppens multipoles.[7]

A chemist inquires more into interaction energies of atoms that are close enough to act strongly on one another. A simple electrostatic solution to the chemist's interaction energy question involves integration in three-dimensional space:[8]

$$E_{\mathrm{es}} \equiv \iiint_{\mathscr{V}} \frac{\rho_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}})\rho_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}})}{|\mathbf{r}_{\mathrm{A}} - \mathbf{r}_{\mathrm{B}}|} \mathrm{d}\mathbf{r}_{\mathrm{A}} \mathrm{d}\mathbf{r}_{\mathrm{B}}. \tag{1.2}$$

This very integral represents the heart of the thesis. It puts together two general charge densities $\big(\rho_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}})$ and $\rho_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}})\big)$ found at two different positions ($\mathbf{r}_{\mathrm{A}}$ and $\mathbf{r}_{\mathrm{B}}$) to work out the electrostatic energy of interaction between them.

A fairly accurate numerical procedure to calculate this core integral takes a long time to complete. The solid core of this thesis encases a method to simplify the numerical three-dimensional integral and thus speed-up the calculation. The method developed in the paper by Nguyen, Kisiel and Volkov, and within the scope of this thesis treats the core integral analytically.[9] Such a mathematical method was transformed into program code to avoid menial computations by hand.

Once this highly relevant program is complete, one can apply it (in principle) to any two molecules whose electron density is represented by a Hansen-Coppens model. One can go further than just any two molecules: parts of the same inorganic complex, two sides of the same molecule, several atoms in close contact, a ligand with an individual amino acid residue from a protein, and so on: a huge variety of chemical and biological systems opens up for exploration.

Bornes to this field of application exist, fences put up by the Hansen-Coppens model itself. Hansen and Coppens worked on it to characterise molecules in crystals. Its biggest application is in high-resolution structure determination from X-ray diffraction experiments on high-quality single crystals. Such crystals are hard to come by, so alternative approaches such as a transfer from a database become appealing. These transfers are indeed possible and databases from both theoretical and experimental sources exist to empower them. The next section gives a brief overview of such databases.

The topics touched upon in this introduction will be expanded in the introductory sections of upcoming chapters, and a more in-depth development of key points will follow those introductions. Before that, a brief detour into similar methods is *de rigueur* to find a place for the method presented in this thesis among its peers.

[6] Buckingham, 1959.
[7] Coppens, 1997, pp. 143–164.
[8] The symbol $\mathscr{V}$ denotes integration in a three-dimensional volume (space).
[9] Nguyen, Kisiel and Volkov, 2018.

1.3 Methods to Calculate Energy

An investigation of intermolecular interactions involves an intelligent electrostatic interaction energy examination. The largest hindrance to such an investigation is the way to obtain a representative model of diffuse charges in interacting molecules with high precision.[10] These approaches normally fall under the density functional theory (DFT).[11]

DFT rests on two pillars: the Hohenberg-Kohn theorems. The first of them states that one needs only the total (nucleus and electron) charge density function $\rho(x, y, z)$ to discover the total energy E of a molecular system in its ground state. This contrasts to most standard quantum theory methods, which demand the use of a wave function $\psi(x_1, y_1, z_1, \chi_1, \ldots x_n, y_n, z_n, \chi_n)$ for a system containing n electrons. This implies the existence of a functional F to calculate the energy $E = F[\rho(x, y, z)]$ from the charge density.

The second Hohenberg-Kohn theorem puts the variational principle into the heart of DFT methodology. Analytically, it points to the global minimum value of the *exact* functional F as the *exact* ground state energy, while the corresponding density represents the exact ground state density. Numerically, one can start with an initial guess $\rho_0(x, y, z)$, calculate the energy E_0, figure out a new density $\rho_1(x, y, z)$ with a smaller energy E_1, and so on iteratively until reaching some convergence criterion. The procedure approximates the *exact* energy and the *exact* density for the investigated molecule's ground state.

This idyllic methodology (especially the point at which one substitutes $4N$ coordinates with only 3) has one major drawback: no one knows the *exact* functional F to calculate the energy $E = F[\rho(x, y, z)]$. It is unclear how to even start looking for such a functional purely from physical principles.

Many have developed ways to untie this Gordian knot. Kohn and Sham have worked with fictitious systems of non-interacting electrons containing a certain amount of electron density that mimic real systems of interacting electrons with the same amount of electron density. A basis set approach analogous to Roothaan's,[12] with basis functions centred on atomic nuclei, pushed DFT in the direction of other quantum chemistry methods at the time.

The use of derivatives ($E = F[\rho, \nabla\rho]$, $E = F[\rho, \nabla\rho, \nabla^2\rho]$ *etc.*) further extended the theory. The advent of hybrid functionals convinced the experimentalists that DFT can address their concerns: such functionals combine energies from DFT with those from quantum chemistry calculations, usually Hartree-Fock (HF), to estimate the total energy.

[10] One can easily be confused by the use of terms like "diffuse charges" and "charge distributions" in the context of charge density analysis. They can refer to both atomic/molecular charges and underlying (nuclear and electron) charge densities, depending on the context. Here they refer to the latter.

[11] One can obtain charge densities from quantum-chemistry calculations that do not fall under the purview of DFT. However, such methods do not focus on the density directly, unlike DFT and the Hansen-Coppens model used in this thesis.

[12] Roothaan was the first to suggest a basis set approach for the Hartree-Fock method, the cornerstone of modern quantum computations.

The downside of DFT calculations is usually not their accuracy, but their computational time. One can not tackle large systems like biomacromolecules with it. An approximation based on DFT or HF, the kernel energy method (KEM), gives much faster results comparable with a full DFT or HF calculation, as the increase in the number of atoms does not affect KEM so much.[13] This method has found successful applications in various biological systems.[14]

An alternative approach tries to model experimental electron densities. This is an arduous procedure: one needs high-quality single crystals, a high-quality X-ray diffractometer to collect data at high-resolution (at least to 0.5 Å) and at low temperature (usually 100 K), and a lot of patience to properly model experimental results. Here one usually uses the aforementioned Hansen-Coppens model, which is based on atom-centred density basis functions (again, an analogy to the Roothaan approach).

The advantage of an experimental approach over a purely theoretical one lies in its correspondence with reality. The *true* functional F would, after applying the second Hohenberg-Kohn theorem, yield the *true* density ρ of the ground state. The *true* functional remains beyond reach, and so does the *true* density from a theoretical perspective. The experimental approach, on the other hand, guarantees that the density is reliable, within a certain precision margin.

The drawback of the approach is that there is no clear pathway from a Hansen-Coppens density to total energy of the investigated system. The reasons are multiple: the scope of an X-ray crystallography experiment is not wide enough (no spin density), the crystal is not always adequate, the experiment takes a long time (even with the most up-to-date technology, the measurement takes at least overnight), the overall model is thus an average across time and across the entire measured crystal *etc.*

The biggest problem, however, lies in the lack of theoretical support: something akin to a functional F is missing to rise from experimental charge density to system energy. One can limit the scope of the problem to electrostatic energy only, and thus to the already mentioned core integral (equation 1.2). This integral pre-supposes that the two densities $\rho_A(\mathbf{r}_A)$ and $\rho_B(\mathbf{r}_B)$ are found at two different locations $\mathbf{r}_A$ and $\mathbf{r}_B$, or else the fraction's denominator equals zero. Hence, the electrostatic energy can only be calculated between two distinct densities ρ_A and ρ_B, making it the *electrostatic interaction energy*.[15] The main topic of this thesis is calculating electrostatic interaction energy when the densities ρ_A and ρ_B come from a Hansen-Coppens model.

Electrostatic potential simplifies the integral 1.2 (as shown in a later chapter), and allows a numerical integration on a three-dimensional grid. This approach, known in literature as exact potential (EP), has been available for the last decade and a half.[16] Standard electrostatic interaction energy

[13] Huang, Massa and Karle, 2005; Mandal *et al.*, 2017.

[14] Massa *et al.*, 2019 discusses many applications of the kernel energy method, and provides references to them.

[15] One can calculate the interaction energy of two densities at the same location, but this is immaterial within the scope of this thesis. The paper from Volkov, King *et al.*, 2006 gives some pointers in this direction for the Hansen-Coppens model.

[16] Volkov, Koritsanszky and Coppens, 2004.

calculations combine numerical exact potential (nEP) at short distance with Buckingham's multipolar moments (MM) approximation at long distance. Programs such as `XDPROP` and `VMoPro` make the nEP/MM method available and easy to apply.[17]

The nEP part of nEP/MM scales poorly for large systems, resulting in diminished performance. To improve calculation speed, Nguyen, Kisiel and Volkov have proposed an analytical method to evaluate the same integral.[18] This thesis follows in their footsteps, implementing the analytical exact potential and multipolar moments method (aEP/MM) with a slight modification. Nguyen, Kisiel and Volkov promise high accuracy with calculation speed improved by two orders of magnitude.

In a newer paper,[19] Nguyen and Volkov further analysed aEP/MM precision and proposed a more precise Fourier-based method, albeit at a certain time cost. Their most recent proposal (with P. Macchi) is a molecular multipolar moments method applicable to crystals, as well as corrections for errors in Ewald summations due to density overlap, following the lead of Bojarowski, Kumar and Dominiak.[20] Benchmark systems for testing their methods consisted of small molecules and oligopeptides with carbon, hydrogen, nitrogen and oxygen atoms.

The previous section briefly addressed the possibility of transferring parameters of a Hansen-Coppens model to chemically similar atoms. This gives an alternative to solving the full (like with DFT) or partial (as with KEM) quantum chemistry problem. One can reconstitute a fairly accurate density model of a biomacromolecular system within seconds, and then apply the full might of the aEP/MM approach to calculate accurate electrostatic interaction energies of various parts of that system in record time. One can find parameter databases with parameters sourced from theoretical (INVARIOM, UBDB - improved with parameters for common ligands) and experimental (ELMAM2) databases.[21]

Another way to tackle large systems is to ignore all the quantum complexity and treat atoms (or even groups of atoms) with the simple laws of Newtonian physics and Coulomb's electrostatic law (equation 1.1). This is the force field approach of molecular mechanics.[22]

Force fields treat atoms as points with a certain mass and a certain charge, decreed by the atom type. Molecules assemble as point masses connected with ideal springs to make "covalent chemical bonds". Force fields treat all intramolecular degrees of freedom (distances, angles and torsion angles) as springs, and typically idealised springs. They also model weak (van der Waals) interactions between non-bonded atoms through a Lennard-Jones or similar potential.[23] They may contain further terms to increase complexity and precision, usually at the cost of generality – the window of application of a force field is often restrained to a certain class of molecules. Furthermore, the parameters in a force

[17] Volkov, Macchi *et al.*, 2016; Jelsch *et al.*, 2005.

[18] Nguyen, Kisiel and Volkov, 2018.

[19] Nguyen and Volkov, 2019.

[20] Bojarowski, Kumar and Dominiak, 2016; Nguyen, Macchi and Volkov, 2020.

[21] Dittrich, Koritsánszky and Luger, 2004 establish INVARIOM, Dominiak *et al.*, 2006 develop UBDB, Kumar, Gruza *et al.*, 2019 expand UBDB with parameters for common ligands, and Domagała *et al.*, 2012 advance ELMAM2.

[22] Newton, 1686, Coulomb, 1785, p. 529, Leach, 2001, pp. 165–252

[23] Lennard-Jones, 1924a; Lennard-Jones, 1924b; Lennard-Jones, 1931.

field are fitted to better reproduce experimental geometries and energies, and they do not have much sense if used outside the context of a force field.

Molecular mechanics can tackle very large systems through the use of force fields. One can also conduct molecular dynamics simulations on the same physical principles, while adding thermal movement to the equation. Such simulations help to clarify experimental results from a molecular perspective.

A fascinating link appears between database transfers for the multipolar model and force fields. The principle for assigning force field parameters by atom types reflects the principle for assigning charge density parameters from a database transfer. The reconstituted charge density model is more elaborate than point charges used in force fields.

One could envisage a set of van der Waals parameters similar to those found in force fields. They would complement the electrostatic energy calculations and attain a better estimate of total interaction energy between two molecules. This topic merits further research outside the scope of this thesis.

Part I

Theoretical Developments

2 To Calculate Electrostatic Energy

2.1 Résumé

Élaborer le calcul d'énergie électrostatique à partir du modèle multipolaire de Hansen et Coppens nécessite une introduction mathématique. La première étape consiste en la définition du modèle de Hansen-Coppens, qui est un modèle inspiré du domaine de la chimie quantique.

Dans le contexte de ce modèle, les molécules ne sont qu'une combinaison linéaire d'atomes, les atomes une combinaison (linéaire) du noyau (une charge ponctuelle) et du nuage électronique, le nuage éléctronique une combinaison linéaire des orbitales, et les orbitales une combinaison linéaire des fonctions de base (équation 2.1). Une section entière est dédiée à la description des coefficients linéaires, des fonctions de base jusqu'aux molécules dans le contexte de l'application prévue. Cette section permet d'oublier toute la complexité d'une molécule pour se focaliser sur la partie fondamentale du calcul : l'intégrale des deux fonctions de base.

La section qui suit aborde la fameuse fonction de Löwdin, le pilier sur lequel repose l'ensemble de la thèse.[1] Elle se base sur le dévéloppement de Nguyen, Kisiel et Volkov pour aboutir à la solution de l'intégrale des deux fonctions de base, chacune decrivant le comportement d'une fraction d'un nuage électronique.[2] La solution nécessite le calcul des fonctions Γ complètes et incomplètes, dont l'approche et les formules sont également abordées.

Certaines questions pratiques sont apparues après l'implementation de ces formules dans un logiciel (cf. le chapitre suivant). Notamment, la rotation des populations multipolaires a posé un obstacle à l'application du calcul sur les multipoles. Il a fallu parcourir la littérature disponible pour recueillir toutes les pièces de ce casse-tête, et elles sont rassemblées dans une section approfondie.

Il est également apparu un problème concernant les atomes lourds, causé par un ordre fâcheux des opérations effectuées. La difficulté a été facilement réglée en établissant le bon ordre des opérations mathématiques.

Finalement, la question des calculs à longue distance se pose. Il est possible d'alléger le temps de calcul dans ce cas en faisant des approximations, dont celles de Buckingham sont les plus courantes.[3] Pour épargner du temps sur le développement du logiciel, le choix a été fait de réutiliser le code existant, en restant fidèle à l'esprit de l'approximation de Buckingham.

[1] Löwdin, 1956.
[2] Nguyen, Kisiel et Volkov, 2018.
[3] Buckingham, 1959.

2.2 Introduction

The previous chapter introduced the notion of electrostatic interaction energy as the central issue of this thesis. It also pointed out that the main model to base those calculations is the Hansen-Coppens multipolar model. This chapter deals with the mathematics behind electrostatic energy calculations.

As mentioned before, diffuse charges within atoms and molecules – electron charges in particular – require integration to obtain reliable results. Description of electron charges considered here rely on J. C. Slater's adapted version of the solution to the Schrödinger equation called Slater radial functions.[4] They find their use in the foundations of many quantum chemistry calculations and models.

The Hansen-Coppens multipolar model, introduced in the previous chapter, has Slater functions in its foundations. High-resolution X-ray crystallography reveals charge density in crystals and molecules that make them up, and the multipolar model depicts such experimental data well. A multipolar model of a particular molecule opens the door to structure exploration through many quantum chemistry techniques, including electrostatic energy calculations.

The way to improve the speed and performance of such calculations (depicted as a timeline of important papers on figure 2.1) is through the use of mathematical tools that simplify the original three-dimensional integral (equation 1.2) to a one-dimensional one. Löwdin proposed the first step: a representation of a displaced Slater radial function. Usually, the origin of a Slater function coincides with the origin of the coordinate system (at $r = 0$), but Löwdin's transformation gives the same function displaced to a point $r = d$ away from the origin. This notion is crucial as it allows two Slater functions to co-inhabit the same coordinate system.[5]

Löwdin introduces the α-function as the Slater function with displaced origin. He uses the Bessel functions to define it rigorously.[6] Jones and Weatherford modify this definition slightly to make them more apt when tackling it with computer code.[7] They work further on the α-function with the goal

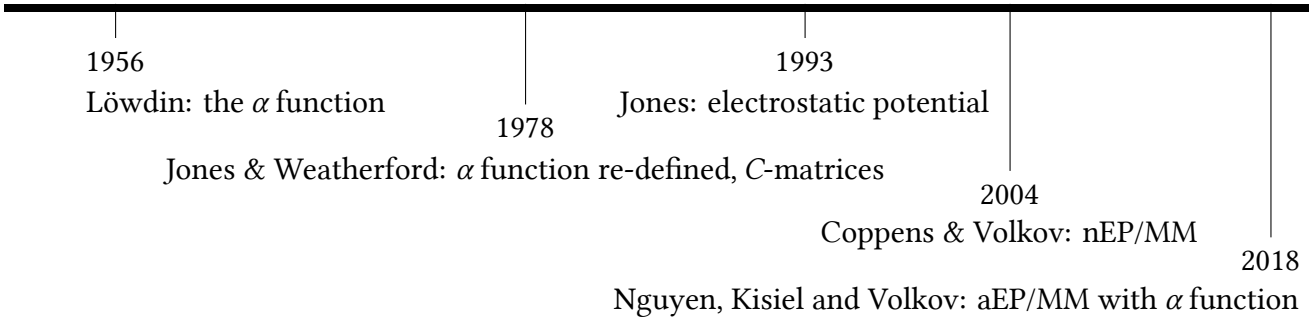


Figure 2.1. Timeline of the most important publications considered within this thesis.

[4] Slater, 1930.
[5] Löwdin, 1956, pp. 96–106.
[6] Ibid.
[7] Jones and Weatherford, 1978.

to apply their results to molecular orbital calculations.[8] Jones also developed a simple formula to estimate the electrostatic potential based on an electron density described as a Slater radial function.[9]

All of these developments happened within the confines of quantum chemistry, with little regard for experimental charge density modelling. The modellers, in particular Volkov, Koritsanszky and Coppens, as well as Fournier *et al.*, developed their own way to calculate electrostatic interaction energies: a numerical integration over a grid.[10] They based their calculations on exact potential and multipolar moments (EP/MM). Volkov, King *et al.* worked on a more robust electrostatic potential based on Γ-functions, a development that attracted some commentary.[11]

The great effort to combine these disparate ideas appeared in a Nguyen, Kisiel and Volkov publication.[12] The authors took the development of Löwdin, Jones and Weatherford on the α-function and applied it to a Slater radial function. Then they used Jones's potential expressed as a simple sum and applied it to another Slater radial function.[13] This second function stayed at the origin.

Integrating the product of an electrostatic potential derived from a Slater function at the origin with the displaced α-function results in a simple integral after many cancellations due to orthonormality of spherical harmonics. The resulting integrals are Γ-functions analogous to those tackled in the Volkov, King *et al.* study.[14] Nguyen, Kisiel and Volkov named their new method *analytical exact potential and multipolar moments* (aEP/MM), and refer to the old one as *numerical* EP/MM (nEP/MM).[15] The main content of this chapter revolves around mathematical details of the developments summarised in this and the previous paragraph.

A few important things remain to be considered from a mathematical standpoint. Firstly, Nguyen, Kisiel and Volkov pass over some important implementation details, like multipole rotation or the necessity of interatomic axis systems. Then, a small improvement came from a change in the order of evaluation of certain operations. This small modification opened the entire periodic system of elements to calculation, at least in principle.

Finally, one must also look into long distance calculations. There reigns a slightly modified solution in the spirit of Buckingham's multipolar moments.[16] These pseudo-multipolar moments (pMM) join the analytical exact potential at short distances to create the main feature of this thesis: the aEP/pMM method for calculating electrostatic interaction energies. A development of cut-offs for unnecessary calculations at long distances also helped to improve computation speed.

Once the mathematical groundwork rests solid, one must search validation through a program implementation to avoid the drudge of repetitive calculations. That is a story for the next chapter.

[8] Jones and Weatherford, 1989; Jones, 1991; Jones, 1992; Jones, 1993.
[9] Jones, 1993.
[10] Volkov, Koritsanszky and Coppens, 2004; Fournier *et al.*, 2009.
[11] Volkov, King *et al.*, 2006; Spackman, 2007; Volkov and Coppens, 2007.
[12] Nguyen, Kisiel and Volkov, 2018.
[13] Jones, 1993.
[14] Volkov, King *et al.*, 2006.
[15] Nguyen, Kisiel and Volkov, 2018.
[16] Buckingham, 1959.

2.3 The Atomic Model

This chapter discusses a particular application of the quantum theory. At its most basic, the quantum theory sees the molecule as a collection of atoms, an atom as a nucleus and an electron cloud, an electron cloud as a collection of orbitals and an orbital as containing electrons (figure 2.2).

From a mathematical standpoint, quantum theory relies heavily on linear algebra. This mathematical basis reflects well the hierarchical view of atoms and molecules. Everything on a higher *échelon* consists of a linear combination of its subordinates: electron density functions[17] χ_{electron} in a linear combination compose an orbital, a linear combination of orbitals ρ_{orbital} gives total electron density ρ_{electron}, and electron density together with a nucleus (holding a positive point charge Z) make up an atom (ρ_{atom}), a linear combination of atoms establishes a molecule. Written in mathematical terms:

$$\begin{aligned}
\rho_{\text{molecule}}(\mathbf{r}) &= \sum_i c_{i,\text{atom}}\, \rho_{i,\text{atom}}(\mathbf{r}_i) \\
&= \sum_i \left[c_{i,\text{atom}}\, Z\delta(\mathbf{r}-\mathbf{r}_i) + c_{i,\text{atom}}\, \rho_{i,\text{electron}}(\mathbf{r}_i) \right] \\
&= \sum_i \left[c_{i,\text{atom}}\, Z\delta(\mathbf{r}-\mathbf{r}_i) + \sum_j c_{i,\text{atom}}\, c_{j,\text{orbital}}\, \rho_{j,\text{orbital}}(\mathbf{r}_i) \right] \\
&= \sum_i \left[c_{i,\text{atom}}\, Z\delta(\mathbf{r}-\mathbf{r}_i) + \sum_j \sum_k c_{i,\text{atom}}\, c_{j,\text{orbital}}\, c_{k,\text{electron}}\, \chi_{k,\text{electron}}(\mathbf{r}_i) \right],
\end{aligned} \tag{2.1}$$

with $\delta(\mathbf{r})$ as the Dirac δ function: $\int_{\mathscr{V}} \delta(\mathbf{r})\, \mathrm{d}\mathbf{r} = 1$. The product of $\delta(\mathbf{r})$ with a charge Z corresponds to the formal definition of a point charge. The linear coefficients c deserve a detailed introduction. The

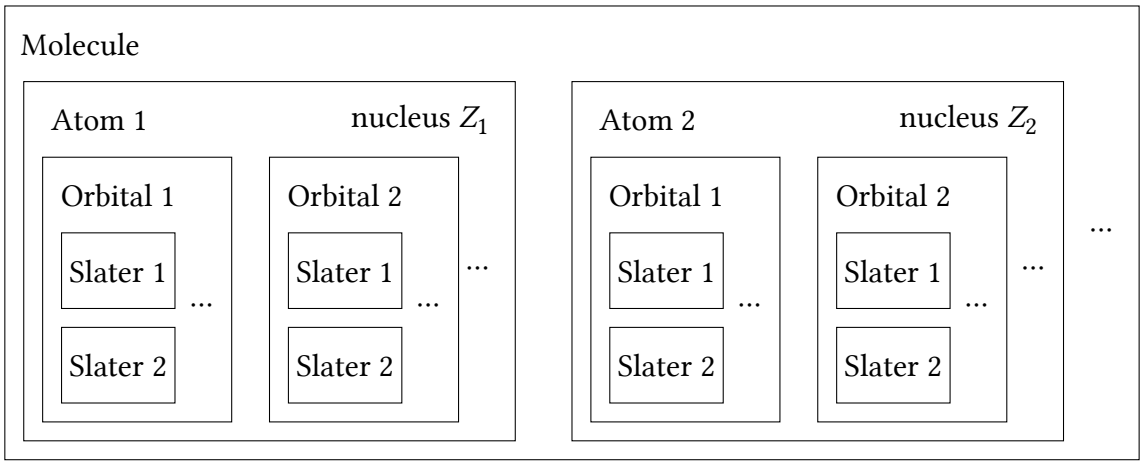


Figure 2.2. The hierarchical model of a molecule: each molecule contains atoms, each atom incorporates a nucleus and an electron cloud composed of orbitals, and each orbital encloses electrons. Electron charge is dispersed, and different functions to mathematically characterise its dispersion exist. Here Slater functions serve that purpose.

[17] The basic electron density functions are denoted as χ_{electron}, to distinguish from (total) electron density ρ_{electron} of an atom. These are not wavefunctions, for which the symbol ψ is exclusively reserved.

indices *i*, *j* and *k* have no general meaning throughout this chapter. They designate the indices of the adjacent summation.

Typically, the coefficients c_{atom} equal 1: a molecule is no more than a simple combination of all its composing atoms. However, crystallographers developed the notions of *occupancy* and *multiplicity* to describe some particular cases they found in some crystal structures.

Crystals consist of highly regular, periodic patterns of repeating motives. The unit cell is the simplest abstract repeating unit of a crystal. Chemically equivalent atoms can take slightly different positions in different unit cells due to imperfections within crystals and the flexibility of certain molecular parts. Occupancy describes how often, on average, an atom is present at a certain position within an average unit cell. Multiplicity describes how often a symmetry element within the crystal creates a symmetrically equivalent atom at the same coordinates as the original one. In general:

$$c_{\text{atom}} = \frac{c_{\text{occupancy}}}{N_{\text{multiplicity}}}, \tag{2.2}$$

but different programs approach multiplicity in different ways.

Further mathematical development demands a choice of a particular atomic model. This thesis hangs on the *Hansen-Coppens multipolar model*, proposed in 1978 to adequately portray empirical electron density from X-ray diffraction experiments on crystals. In it, the electron density of an atom takes the following outline:[18]

$$\rho_{\text{atom}}(\mathbf{r}) = P_{\text{core}}\rho_{\text{core}}(r) + P_{\text{val}}\rho_{\text{valence}}(\kappa r) + \sum_{l=0}^{l_{\max}} R_l(\kappa_l' r) \sum_{m=-l}^{l} P_{lm} d_{lm}(\theta, \phi). \tag{2.3}$$

The first two terms represent core and valence electron densities (ρ_{core} and ρ_{val}). These normally come from *in vacuo* calculations on individual atoms and are moulded to be spherical and normalised to one electron.[19] Each core orbital *i* contains $N_{i,\text{core}}$ electrons, so they all contain $N_{\text{core,total}} = \sum_i N_{i,\text{core}}$ electrons. The total core density $\rho_{\text{core}}(r)$ consists of core orbital densities in proportion to their neutral populations:

$$\rho_{\text{core}}(r) = \sum_i \frac{N_{i,\text{core}}}{N_{\text{core,total}}} \rho_{i,\text{orbital,core}}(r) \tag{2.4}$$

The core density expressed in these terms represents a density of one electron partitioned through all core orbitals. Real atoms contain P_{core} electrons (called the *core population*). One could use this as a refinement parameter, but the almost universal approach is to fix $P_{\text{core}} = N_{\text{core,total}}$. This effectively cancels the denominator of the previous equation.

[18] Hansen and Coppens, 1978. Please note that the κ^3 and κ'^3 terms are not missing from this equation. They are a consequence of radial function normalisation, and thus show up in the radial normalisation constant (equation 2.11) in this detailed treatement. The model equation with these constants is commonplace and instructive to explain the meaning of κ and κ' terms for those interested in using refinement software to obtain a good charge density model. Working with Slater radial functions directly, however, profits from appointing these terms to their place of origin.

[19] Clementi and Roetti, 1974; Koga, Tatewaki and Thakkar, 1993; Koga, Watanabe *et al.*, 1995; Koga, Kanayama *et al.*, 2000.

Valence orbitals within the Hansen-Coppens model behave in analogous ways. A valence orbital i contains $N_{i,\mathrm{val}}$ electrons, and they all contain $N_{\mathrm{val,total}} = \sum_i N_{i,\mathrm{val}}$ electrons. They can contract or expand relative to their "standard" size by adjusting the parameter κ. The expression for a total valence density normalised to one electron looks like this:

$$\rho_{\mathrm{valence}}(\kappa r) = \sum_i \frac{N_{i,\mathrm{val}}}{N_{\mathrm{val,total}}} \rho_{\mathrm{orbital,valence}}(\kappa r) \tag{2.5}$$

Atoms contain P_{val} electrons in their valence orbitals (known as their *valence population*). Unlike core populations, one often refines this parameter to determine how many electrons there really are around an atom – in other words, how much extra or how much fewer charge does that atom have. Valence orbitals share this extra charge, or the charge depletion, in proportion to their neutral populations $N_{i,\mathrm{val}}$.

The final sum in the Hansen-Coppens model (equation 2.3) is the sum of terms that represent electron density multipoles, giving the model its alternative name: *the multipolar model.* The terms P_{lm} are the multipolar coefficients or *multipolar populations* attributed to each individual atom. This sum exists outside of the traditional orbital structure. As a summary, the orbital linear coefficient has the form:

$$c_{\mathrm{orbital}} = \begin{cases} N_{\mathrm{core}} & \text{for core orbitals,} \\ \frac{N_{\mathrm{val}}}{N_{\mathrm{val,total}}} P_{\mathrm{val}} & \text{for valence orbitals,} \\ P_{lm} & \text{for multipoles.} \end{cases} \tag{2.6}$$

Orbital densities for core and valence orbitals come from quantum calculations, so one calculates orbital densities as a square of the wave function:

$$\rho_{\mathrm{orbital}}(r) = \psi^2(r). \tag{2.7}$$

These densities have an alternative definition using all three spherical coordinates:

$$\rho_{\mathrm{orbital}}(\mathbf{r}) = \psi^2(r) d_{00}(\theta, \phi). \tag{2.8}$$

The two definitions are equivalent since $d_{00}(\theta, \varphi)$ is spherical. The calculations discussed in this chapter require the explicit use of the $d_{00}(\theta, \varphi)$ function (defined below) for orbital densities. The wave functions used to reconstruct orbital density depend only on the distance from origin and has the form of a simple sum:

$$\psi(r) = \sum_i R_i(r). \tag{2.9}$$

Both the orbital wave functions (equation 2.9) and the multipolar terms of the Hansen-Coppens model (equation 2.3) depend on the same kind of radial function $R(r)$. The particular kind used commonly for experimental charge density models is the Slater function:

$$R(r) = N_{\mathrm{r}} c r^n e^{-\zeta r}, \tag{2.10}$$

with n (an integer), c and ζ (real numbers) as coefficients coming from *in vacuo* calculations[20] (for orbitals) or empirical observation (for multipoles; $c = 1$ in that case). The normalisation coefficient N_{r}, to normalise the radial part, takes the form of:

$$N_{\mathrm{r}} = \kappa^3 (2\zeta)^n \sqrt{\frac{2\zeta}{(2n)!}} \tag{2.11}$$

Finally, a further adjustable parameter κ (a real number), proposed by Hansen and Coppens, serves to improve the fit over experimental data.

An important property of the radial function has to do with multiplication. A product of two different radial functions $R_{\mathrm{A}}(r)$ and $R_{\mathrm{B}}(r)$ is a function of the same type:

$$R_{\mathrm{A}}(r) R_{\mathrm{B}}(r) = N_{\mathrm{A,r}} c_{\mathrm{A}} r^{n_{\mathrm{A}}} e^{-\zeta_{\mathrm{A}} \kappa_{\mathrm{A}} r} \cdot N_{\mathrm{B,r}} c_{\mathrm{B}} r^{n_{\mathrm{B}}} e^{-\zeta_{\mathrm{B}} \kappa_{\mathrm{B}} r} = N_{\mathrm{AB,r}} c_{\mathrm{AB}} r^{n_{\mathrm{AB}}} e^{-\zeta_{\mathrm{AB}} \kappa_{\mathrm{AB}} r} = R_{\mathrm{AB}}(r). \tag{2.12}$$

Since the wave function (equation 2.9) is a linear combination of radial functions (equation 2.10), its square also becomes a linear combination of radial functions (equation 2.10). It could have been defined like this from the beginning, but the equation 2.9 reflects a necessary step to get from coefficients available in literature to a full description of the density of a spherical orbital. The spherical orbital density (equation 2.8) is thus, in the most general terms, a linear combination of products of a radial function with a $d_{00}(\theta, \varphi)$ spherical harmonic.

Multipoles (equation 2.3) and orbital densities (equation 2.8) depend on *spherical harmonic* functions $d_{lm}(\theta, \varphi)$. Spherical harmonics are widely used functions, and have different normalisations to adapt to different applications. Density-normalised real spherical harmonics $d_{lm}(\theta, \varphi)$ are the convention within the field of experimental electron density. They take the following form:

$$d_{lm}(\theta, \varphi) = \begin{cases} N_{\mathrm{a}} P_l^m(\cos\theta) \cos(m\varphi) & m \geq 0 \\ N_{\mathrm{a}} P_l^{|m|}(\cos\theta) \sin(|m|\varphi) & m < 0, \end{cases} \tag{2.13}$$

with associated Legendre polynomials[21] $P_l^{|m|}(x)$ taking $x = \cos\theta$ as an argument. Consider the three first-order spherical harmonics ($l = 1$) as an example:[22]

[20] Clementi and Roetti, 1974 made the first extensive tables of spherically-averaged electron densities around atoms in the vacuum (up to xenon, $Z = 54$, *i.e.* the first five rows). More recently, Koga, Tatewaki and Thakkar, 1993 and Koga, Watanabe *et al.*, 1995 improved on this work. Koga, Kanayama *et al.*, 2000 expand the tables to incorporate much of the periodic system (up to lawrencium, $Z = 103$).

[21] Hans-Jurgen Weber *et al.*, 2004.

[22] The associated Legendre polynomials $P_1^0(x) = x$ and $P_1^1(x) = -\sqrt{1 - x^2}$. The second formula simplifies due to the trigonometric identity $\sin(\theta) = \sqrt{1 - \cos^2(\theta)}$.

$$d_{1m}(\theta,\varphi) = \begin{cases} -N_\mathrm{a}\sin(\theta)\cos(\varphi) & m = 1 \\ N_\mathrm{a}\cos(\theta) & m = 0 \\ -N_\mathrm{a}\sin(\theta)\sin(\varphi) & m = -1, \end{cases}$$

The angular normalisation coefficient has the following form:

$$N_\mathrm{a} = (-1)^m \sqrt{\frac{4\pi}{(2l+1)}\frac{(l+|m|)!}{(l-|m|)!}} l_0 \left[\int_0^\pi \int_0^{2\pi} \left(\frac{|d_{lm}(\theta,\varphi)|}{N_\mathrm{a}}\right) \sin\theta \,\mathrm{d}\varphi \,\mathrm{d}\theta\right]^{-1}. \tag{2.14}$$

The coefficient l_0 is 1 when $l = 0$, and 2 otherwise. Michael and Volkov have computed the integrals from equation 2.14 to a 35-digit precision.[23] These values exist as constants in the source code developed within the context of this thesis.

Under the normalisation from equation 2.14, the $d_{00}(\theta,\varphi)$ harmonic function multiplied with an appropriate radial function encloses a sphere containing the electron density of one electron, while all other spherical harmonics represent the transfer of one electron from their negative to their positive lobes.

Spherical harmonics are mutually orthogonal. Integrating a product of two spherical harmonics A and B yields zero, except when $l_\mathrm{A} = l_\mathrm{B}$ and $m_\mathrm{A} = m_\mathrm{B}$. Exploiting this property reduces significantly the number of necessary integrations further down the line.

The final linear coefficient (equation 2.1) for an electron function is thus:

$$N_\mathrm{electron} = N_\mathrm{r} N_\mathrm{a} c \tag{2.15}$$

The basic electron density function thus becomes, in the most general terms:

$$\chi_\mathrm{electron}(\mathbf{r}) = R(r) d_{lm}(\theta,\varphi), \tag{2.16}$$

using definitions from equations 2.10 and 2.13. This definition is applicable to spherical orbitals since they are linear combinations of products of a radial function multiplied by the $d_{lm}(\theta,\varphi)$ spherical harmonic (equations 2.8, 2.9 and 2.12). One must multiply them by the appropriate coefficients defined in equations 2.2, 2.6 and 2.15 to build an electron cloud from them.

This development allows one to put aside all the complications of an atom and focus on the most basic components: one basic electron density function $\chi(\mathbf{r})$ (to assemble the electron cloud) and one point charge $Z\delta(\mathbf{r})$ (as a nucleus). The attention in the next chapter turns towards them.

[23] Michael and Volkov, 2015.

2.4 THE COULOMB INTEGRAL

The core subject of this thesis is how to evaluate the Coulomb integral for a pair of atoms **A** and **B** described through the Hansen-Coppens multipolar model. The Coulomb integral expresses the core idea of electrostatics: how strongly two static charges attract or repel each other. Defined in the most general terms, it takes the following form (reproduced from *General Introduction*, and using atomic units):

$$E_{\text{es}} \equiv \iint_{\mathscr{V}} \frac{\rho_{\text{A}}(\mathbf{r}_{\text{A}})\rho_{\text{B}}(\mathbf{r}_{\text{B}})}{|\mathbf{r}_{\text{A}} - \mathbf{r}_{\text{B}}|} \mathrm{d}\mathbf{r}_{\text{A}} \mathrm{d}\mathbf{r}_{\text{B}}. \tag{1.2}$$

2.4.1 Nucleus-Nucleus and Nucleus-Electron Terms

The two atoms each have a point-charge nucleus at origin with a charge Z, formally defined as $\rho(\mathbf{r}) = Z\delta(\mathbf{r})$. Integrating the Dirac δ function yields 1, so the **n**ucleus-**n**ucleus term is simple:[24]

$$E_{\text{es,nn}} = \frac{Z_{\text{A}} Z_{\text{B}}}{d}, \tag{2.17}$$

with d denoting the distance between the two nuclei.

The second term to consider involves the attraction between a nucleus of atom A and the electron cloud of atom B (and also with the order reversed). Instead of taking the whole electron cloud at once, one should consider its simplest element, a basic function from equation 2.16. Taking only the electron part at first gives a new definition:

$$\Phi(\mathbf{r}) \equiv \int_{\mathscr{V}} \frac{\chi(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \mathrm{d}\mathbf{r}'. \tag{2.18}$$

This is the electrostatic potential (ESP): the effort needed to bring a positive elementary charge $+e$ from infinite distance to the point $\mathbf{r}$.[25] The electrostatic energy between the **n**ucleus of atom A and one **e**lectron function of the atom B simplifies to a multiplication:

$$E_{\text{es,en}} = Z_{\text{B}} \int_{\mathscr{V}} \frac{\chi_{\text{A}}(\mathbf{r}_{\text{A}})}{|\mathbf{r}_{\text{A}} - \mathbf{r}_{\text{B}}|} \mathrm{d}\mathbf{r}_{\text{A}} = Z_{\text{B}} \Phi_{\text{A}}(\mathbf{r}_{\text{B}}) \tag{2.19}$$

In this particular case, the electrostatic potential emanating from the electron cloud of atom A must be evaluated at the position of the atom B to calculate its influence on the nucleus B. This term is asymmetrical, so it must be computed twice: once for the electron cloud of the atom A and the nucleus of the atom B (as described), and once more for the electron cloud of the atom B and the nucleus of the atom A (with labels switched).

[24] Coulomb, 1785, p. 529.
[25] Griffiths, 2021, pp. 113–166.

There is a large literature on deriving electrostatic potential from electron density functions like the one in equation 2.16.[26] Jones substitutes the integral with a three-term sum:

$$\Phi(\mathbf{r}) = \sqrt{\frac{4\pi}{2l+1}}\left[\phi_1(r) + \phi_2(r) + \phi_3(r)\right] d_{lm}(\theta, \varphi), \tag{2.20}$$

The individual terms within the Jones potential have the following form:

$$\phi_1(\kappa, r) = \frac{\nu!}{\zeta^{\nu+1}}(\kappa r)^{-l-1}, \tag{2.20a}$$

$$\phi_2(\kappa, r) = -e^{-\zeta\kappa r}\sum_{k=0}^{\nu}\frac{\nu!}{(\nu-k)!\zeta^{k+1}}(\kappa r)^{\nu-k-l-1}, \tag{2.20b}$$

$$\phi_3(\kappa, r) = e^{-\zeta\kappa r}\sum_{k=0}^{\mu}\frac{\mu!}{(\mu-k)!\zeta^{k+1}}(\kappa r)^{\mu-k+l}. \tag{2.20c}$$

with $\nu = n+l+2$ and $\mu = n-l+1$.[27] Calculating this potential in a computer program is straightforward.

2.4.2 Electron-Electron Term

The final term of interest is between an **e**lectron cloud and its counterpart **e**lectron cloud. Again the focus shifts from the general density to the most basic functions:

$$E_{\text{es,ee}} = \iint_{\mathscr{V}} \frac{\chi_{\text{A}}(\mathbf{r}_{\text{A}})\chi_{\text{B}}(\mathbf{r}_{\text{B}})}{|\mathbf{r}_{\text{A}} - \mathbf{r}_{\text{B}}|}\mathrm{d}\mathbf{r}_{\text{A}}\mathrm{d}\mathbf{r}_{\text{B}} \tag{2.21}$$

Substituting the electrostatic potential (equation 2.18) results in an equation easier to tackle. The Jones potential (equation 2.20) helps in particular to alleviate the integration with a series of sums:

$$E_{\text{es,ee}} = \int_{\mathscr{V}} \Phi_{\text{A}}(\mathbf{r}_{\text{B}})\chi_{\text{B}}(\mathbf{r}_{\text{B}})\,\mathrm{d}\mathbf{r}_{\text{B}}. \tag{2.22}$$

Behing the curtain of this equation rest the two radius-vectors $\mathbf{r}_{\text{A}}$ and $\mathbf{r}_{\text{B}}$ with two different origins. Such a dependence prompts one to consider numerical integration. It paints a clear path to application: calculate the ESP of one atom and the density of another at the points of a three-dimensional grid surrounding both atoms, multiply them together and sum up the products of that multiplication. This procedure assures numerical stability, paid through computational time required to calculate values of ESP, density and their product on 3D grids.

[26] One can find the definition of the electrostatic potential in classical electromagnetism textbooks, such as Griffiths, 2021, pp. 113–166. The paper Jones, 1993 develops the electrostatic potential on the basis of Slater radial functions, drawing heavily from the work of Wahl, Cade and Roothaan, 1964. The chapter 8 of Coppens, 1997, pp. 165–191, brings a detailed introduction into this topic, summarising many of the mentioned papers. More recently, Volkov, King *et al.*, 2006 developed a more robust electrostatic potential based on Γ-functions. On a similar topic, the paper Johnson *et al.*, 1993 describes an algorithm applicable to both Slater and Gaussian radial potentials; the latter are more commonly used in quantum chemistry calculations.

[27] Jones, 1993, p. 23; Nguyen, Kisiel and Volkov, 2018, p. 528.

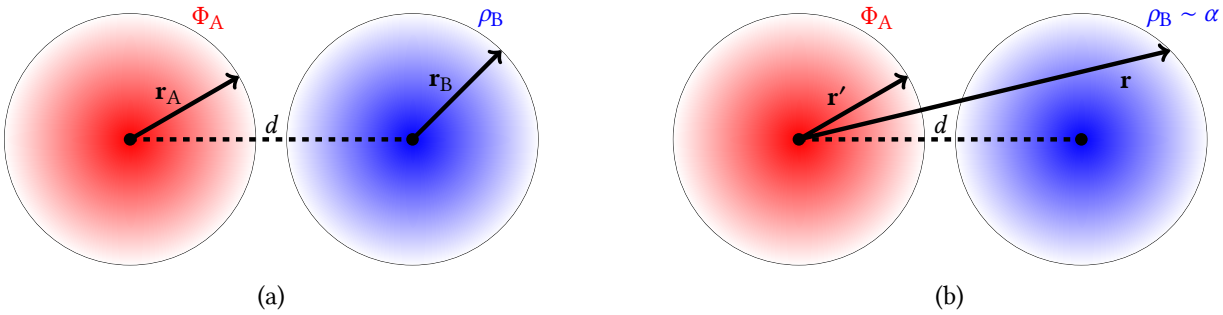


Figure 2.3. Illustration of the Löwdin's α-transformation. (a) Before: the potential Φ and density ρ functions are in their own co-ordinate ($\mathbf{r}_A$ and $\mathbf{r}_B$), whose origins are separated by a distance d. (b) After: Both functions now occupy the same co-ordinate system with the same origin ($\mathbf{r}$ and $\mathbf{r}'$ for consistency with equation 2.18), since ρ is now re-written in terms of Löwdin's α-function.

An analytical solution to this integral would result in a faster computation if a 1D integral could replace 3D grids. Such an approach remains hopeless so long as the ESP and density functions depend on different radius-vectors $\mathbf{r}_A$ and $\mathbf{r}_B$.

Löwdin's α-function

Löwdin's α-function helps to displace a Slater function from the origin of its coordinate system. It moves the origin of a Slater function to a point $d = |\mathbf{r}_B - \mathbf{r}_A|$ away from the origin of the coordinate system. The basis function χ, seen from this new perspective, takes shape through the use of the α-function:

$$\begin{aligned} \chi(\mathbf{r}_B) &= R(r_B) d_{l_B m_B}(\theta_B, \varphi_B) \rightarrow \\ \chi(\mathbf{r}) &= \sum_{l'=0}^{\infty} N_{l'} \alpha_{l'}^{n'_B l_B m_B}(d, r) d_{l' m_B}(\theta, \varphi), \end{aligned} \tag{2.23}$$

with $n'_B = n_B + 1$ and a normalisation coefficient:

$$N_{l'} = \frac{1}{\zeta_B^{n_B}} \sqrt{\frac{(2l_B+1)}{(2l'+1)} \frac{(l_B+m_B)!}{(l_B-m_B)!} \frac{(l'-m_B)!}{(l'+m_B)!}}.$$

This coefficient replaces the N_A coefficient (equation 2.14). Figure 2.3 depicts this transformation.[28]

Both functions now co-inhabit the same coordinate system. The electrostatic energy integral (equation 2.22) takes the following form in this new system:

$$E_{es,ee} = \int_{\mathscr{V}} \Phi_A(\mathbf{r}) \chi_B(\mathbf{r}) \, d\mathbf{r}. \tag{2.24}$$

An analytical solution of this integral exists. The description thus far suffices to understand the α-function method on a conceptual level.

To proceed into details required mainly for a code implementation, one needs a proper definition of Löwdin's α-function:

[28] Löwdin, 1956, pp. 96–106.

$$\alpha_{l'}^{nlm}(d,r) = \frac{(2l'+1)}{2}\frac{(l'-m)!}{(l'+m)!}\sum_{i=0}^{n+l+l'}\sum_{j=0}^{n+l'} C_{l'}^{nlm}(i,j)H_{ij}(r)(\zeta\kappa d)^{i-l-l'-1}(\zeta\kappa r)^{i-l'-1}, \tag{2.25}$$

with:

$$H_{ij}(r) = \begin{cases} e^{-\zeta\kappa d}\left[(-1)^j e^{\zeta\kappa r} - e^{-\zeta\kappa r}\right] & r < d, \\ e^{-\zeta\kappa r}\left[(-1)^i e^{\zeta\kappa d} - e^{-\zeta\kappa d}\right] & r > d. \end{cases} \tag{2.26}$$

One can pre-compute the elements of the C-matrix $[C_{l'}^{nlm}(i,j)]$ using this formula:

$$\begin{aligned}\sum_{i=0}^{n+l-|m|+l'}\sum_{j=0}^{n+l'} C_{l'}^{nlm}(i,j)d^i r^j = &\sum_{p=0}^{\frac{l+m}{2}}\sum_{q=0}^{l+m-2p}\sum_{\nu=0}^{l+m-2p-q}\sum_{p'=0}^{\frac{l'-m}{2}}\sum_{q'=0}^{l'-m-2p'}\sum_{\nu'=0}^{l'-m-2p'-q'}\sum_{k=0}^{t}\sum_{k'=0}^{t-k} \\ &\frac{z^{n+l+2l'-2p'-2\nu'-2\nu-k-k'}d^{2p'+2\nu+2\nu'+k'}(-1)^{\nu+q'+p+p'+l}}{4^{l+l'+p-p'}(l-p)!(l'-p')!p!p'!q!q'!\nu!\nu'!k'!} \\ &\cdot\frac{(2l-2p)!(2l'-2p')t!}{(l+m-2p-q-\nu)!(l'-m-2p'-q'-\nu')!(t-k-k')!}\end{aligned} \tag{2.27}$$

with $t = n-l+2p+2q+2q'$. The matrix elements are coefficients in the polynomial obtained from the right-side expression, and their indices correspond to exponents of d and r in the same polynomial. The only relevant C-matrices have $n \geq l \geq l' \geq m$.

Here is an example of a C-matrix to clarify the use of equation 2.27. Let $n = l' = l = m = 1$. The evaluation of the right-hand side yields the expression $\sum_{i=0}^{2}\sum_{j=0}^{2} C_1^{111}(i,j)d^i r^j = -3 - 3d - d^2 - 3r - 3rd - rd^2 - r^2 d - r^2 d^2$. The value -3 goes to matrix positions (0,0), (0,1), (1,0), (1,1); positions (0,2), (1,2), (2,0), (2,1) take -1; and the remaining position (2,2) is 0. This makes the matrix:

$$\mathbf{C}_1^{111} = \begin{bmatrix} -3 & -3 & -1 \\ -3 & -3 & -1 \\ -1 & -1 & 0 \end{bmatrix}.$$

One can check the correctness of C-matrices from observed properties[29] or by comparing them to those known from literature sources.[30]

[29] Some properties of C-matrices: (1) The dimension of a $C_{l'}^{nlm}$-matrix is $n+l-m+l'+1$ by $n+l'+1$; (2) $C_0^{000} = [1]$; the last element of all other C-matrices is 0; (3) if a C-matrix is square, it is symmetrical; (4) elements of a C-matrix are zero or of the same sign as $(-1)^{l+m-l'}$; (5) C-matrices for any integer n and $l' = l = m = 0$ have their $(0, i)$ and $(i, 0)$-th elements equal to $n!/i!$.

[30] Jones and Weatherford, 1978, p. 487,Jones and Weatherford, 1989, p. 236

Electron-Electron Integral

With the Löwdin α-function in hand, the potential Φ_A (equation 2.20) and the density χ_B (equation 2.16) re-defined in the context of the α-function (equation 2.23) coalesce within the transformed electrostatic integral (equation 2.24). Nguyen, Kisiel and Volkov give the full derivation, whose final (but not general) solution is as follows:[31]

$$\begin{aligned} E_{ee}(\kappa_A, \kappa_B, \mathbf{r}) = k_0 \sum_{i=0}^{i_{max}} \sum_{j=0}^{j_{max}} C_{l_A}^{n'_B l_B m_B}(i,j)(\kappa_B \zeta_B)^{i+j} d_B^i \\ \cdot \left(\nu! \frac{\kappa_A^{-l_A-1}}{\zeta_A^{\nu+1}} I_1 - \sum_{k=0}^{\nu} \frac{\nu!}{(\nu-k)!} \frac{\kappa_A^{\nu-k-l_A-1}}{\zeta_A^{k+1}} I_2 + \sum_{k=0}^{\mu} \frac{\mu!}{(\mu-k)!} \frac{\kappa_A^{\mu-k+l_A}}{\zeta_A^{k+1}} I_3 \right) \end{aligned} \tag{2.28}$$

with the sum limits $i_{max} = n_B + l_B - |m_B| + l_A + 1$, $j_{max} = n_B + l_A + 1$, $\nu = n_A + l_A + 2$, and $\mu = n_A - l_A + 1$; normalisations N_r and N_a from equations 2.11 and 2.14 respectively; for notational convenience, $n'_B = n_B + 1$. The remaining undefined terms are:

$$k_0 = \frac{2\pi}{d^{l_B+l_A+1} \zeta_B^{n_B+l_B+2l_A+2} \kappa_B^{l_B+2l_A+2}} \cdot \sqrt{\frac{2l_B+1}{2l_A+1} \frac{(l_B+m_B)!}{(l_B-m_B)!} \frac{(l_A-m_B)!}{(l_A+m_B)!}}, \tag{2.28a}$$

$$I_1 = (-1)^j \epsilon_B^{-1}(d) \int_0^d r_1 \epsilon_B(r) \mathrm{d}r + (-1)^i \epsilon_B(d) \int_d^\infty r_1 \epsilon_B^{-1}(r) \mathrm{d}r - \epsilon_B^{-1}(d) \int_0^\infty r_1 \epsilon_B^{-1}(r) \mathrm{d}r, \tag{2.28b}$$

$$I_2 = (-1)^j \epsilon_B^{-1}(d) \int_0^d r_2 \epsilon_{A-B} \mathrm{d}r + (-1)^i \epsilon_B(d) \int_d^\infty r_2 \epsilon_{A+B} \mathrm{d}r - \epsilon_B^{-1}(d) \int_0^\infty r_2 \epsilon_{A+B} \mathrm{d}r, \tag{2.28c}$$

$$I_3 = (-1)^j \epsilon_B^{-1}(d) \int_0^d r_3 \epsilon_{A-B} \mathrm{d}r + (-1)^i \epsilon_B(d) \int_d^\infty r_3 \epsilon_{A+B} \mathrm{d}r - \epsilon_B^{-1}(d) \int_0^\infty r_3 \epsilon_{A+B} \mathrm{d}r, \tag{2.28d}$$

with the following substitutions:

$$r_1 = r^{j-2l_A} \tag{2.29a}$$

$$r_2 = r^{j-2l_A+n-k} \tag{2.29b}$$

$$r_3 = r^{j+1+m-k} \tag{2.29c}$$

[31] Nguyen, Kisiel and Volkov, 2018, pp. 528–529.

$$\epsilon_{\mathrm{B}}(r) = e^{\kappa_{\mathrm{B}}\zeta_{\mathrm{B}}r} \tag{2.29d}$$

$$\epsilon_{\mathrm{A}\pm\mathrm{B}} = e^{-(\kappa_{\mathrm{A}}\zeta_{\mathrm{A}}\pm\kappa_{\mathrm{B}}\zeta_{\mathrm{B}})r} \tag{2.29e}$$

The following equations are useful to evaluate these integrals:

$$\int_0^\infty r^n e^{-br}\mathrm{d}r = \begin{cases} \frac{n!}{b^{n+1}} & n \geq 0 \\ \text{set to } 0 & n = -1 \\ \sum_{i=1}^{|n|-1}(-b)^{|n|-1}\frac{(|n|-i-1)!}{(|n|-1)!(|n|-i)!} & n \leq -2 \end{cases} \tag{2.30a}$$

$$\int_a^\infty r^n e^{-br}\mathrm{d}r = \begin{cases} \frac{e^{-ab}}{b^{n+1}}\sum_{i=0}^{n}\frac{n!}{(n-i)!}(ab)^{n-i} & n \geq 0 \\ \text{set to } 0 & n = -1 \\ -(-b)^{|n|-1}\sum_{i=1}^{|n|-1}\frac{-e^{-ab}(|n|-i-1)!}{(-ab)^{|n|-i}(|n|-1)!} & n \leq -2 \end{cases} \tag{2.30b}$$

$$\int_0^a r^n e^{-br}\mathrm{d}r = \begin{cases} \int_0^\infty r^n e^{-br}\mathrm{d}r - \int_a^\infty r^n e^{-br}\mathrm{d}r & n < p_{\mathrm{lim}}, \\ \mu(ab, n+1)e^{-ab+(n+1)\ln a} & \text{otherwise.} \end{cases} \tag{2.30c}$$

The last integral has several possible complications. Firstly, it can occasionally have $b = 0$, in which case it boils down to a much simpler integral:

$$\int_0^a r^n\mathrm{d}r = \begin{cases} \frac{a^{n+1}}{n+1} & n \neq -1 \\ \text{set to } 0 & n = -1 \end{cases} \tag{2.30d}$$

Secondly, when n is large and not negative, and the product ab small, subtraction (the standard case in equation 2.30c) results in serious precision loss. The alternative formula requires the calculation of a continued fraction:

$$\mu(ab, n+1) = \frac{\alpha_1}{\beta_1+}\frac{\alpha_2}{\beta_2+}\frac{\alpha_3}{\beta_3+}\cdots = \frac{\alpha_1}{\beta_1 + \frac{\alpha_2}{\beta_2+\frac{\alpha_3}{\beta_3+\cdots}}}. \tag{2.30e}$$

To evaluate the fraction in the particular case of interest, one must set $\alpha_1 = 1$; for other values of m, $\alpha_{2m} = -(m+n)ab$ and $\alpha_{2m+1} = mab$; for all m, $\beta_m = m+n$. Algorithm recipe books supply algorithms to evaluate continued fractions until desired convergence. This evaluation is more time-consuming than calculating the difference of the other two integrals (the standard calculation in 2.30c, using 2.30a and 2.30b), and does not fit well to all domains of interest. It is therefore limited only to non-negative values of n larger than or equal to the following value:

$$p_{\mathrm{lim}}(ab) = \begin{cases} 5\sqrt{|ab|} - 5 & ab < -9, \\ 0 & -9 \leq ab \leq 0, \\ ab & ab > 0. \end{cases} \tag{2.30f}$$

Role of Spherical Harmonics

This chronicles all things necessary for the calculation of interaction energy between two Slater functions representing diffuse electron density around a nucleus. Spherical harmonics have mysteriously vanished in this narration (absent in equation 2.28). Löwdin's α-transformation (equation 2.23) has an infinite sum of spherical harmonics $d_{l'm_\mathrm{B}}(\theta,\varphi)$. They multiply with their counterparts $d_{l_\mathrm{A}m_\mathrm{A}}(\theta,\varphi)$ from electrostatic potential (equation 2.20). The integral of this product is:

$$\sum_{l'=0}^{\infty}\int_0^{\pi}\int_0^{2\pi} d_{l_\mathrm{A}m_\mathrm{A}}(\theta,\varphi)d_{l'm_\mathrm{B}}(\theta,\varphi)\,\mathrm{d}\varphi\,\mathrm{d}\theta = \sum_{l'=0}^{\infty}\delta_{l_\mathrm{A}l'}\delta_{m_\mathrm{A}m_\mathrm{B}}. \tag{2.31}$$

Because Kronecker's δ_{ij} is one when $i = j$ and zero otherwise, the equation 2.28 contains l_A in place of l'. The main expression (equation 2.28) always evaluates to zero if $m_\mathrm{A} \neq m_\mathrm{B}$. When $m_\mathrm{A} = m_\mathrm{B}$ and $l_\mathrm{A} \neq l_\mathrm{B}$, it does not have to equal zero. This is a direct consequence of Löwdin's α-transformation. This means in practical terms that an interaction energy can be calculated between, for example, a dipole A and a quadrupole B, provided that $m_\mathrm{A} = m_\mathrm{B}$.

The solution given in the equation 2.28 is final, but not general. The Löwdin α-function introduces a pre-supposition: the spherical harmonics inhabit the interatomic coordinate system. Such a system has the axis z pointing from the centre of the atom A to the centre of the atom B. The other two axes are arbitrary, bounded only to make an orthogonal basis. The equation 2.28 yields correct results only within the context of an interatomic coordinate system. Otherwise, it misapprehends the multipole density.

Löwdin did not discuss multipole densities as he dealt largely with spherical densities, which do not depend on the axis system in use. The early work of Jones and Weatherford continues in that direction,[32] but they mention in their later paper that the displacement is specifically along the z-axis.[33] Nguyen, Kisiel and Volkov repeat this, thus implying the necessity of an interatomic coordinate system in their paper.[34]

As a side note, it is possible to calculate the interaction B·A with the same interatomic coordinate system as for the A·B calculation. One must take care to correct the sign of the end result when l_A and l_B are both even or both odd. Paying attention to this avoids some invalid integrals when $l_\mathrm{B} \geq l_\mathrm{A} + 3$ and saves computational time on re-calculating the interatomic coordinate system.

[32] Jones and Weatherford, 1978.
[33] Jones and Weatherford, 1989.
[34] Nguyen, Kisiel and Volkov, 2018.

Final Thoughts

To calculate the electrostatic energy between molecules, one typically sums the contributions of electrostatic energies involving all atom pairs from the two molecules. Interactions between atoms A and B in a pair further separate into nucleus-nucleus, nucleus-electron and electron-electron contributions:

$$E_{\text{es,total}} = E_{\text{es,nn}} - E_{\text{es,en}} + E_{\text{es,ee}}. \tag{2.32}$$

This chapter so far detailed how to obtain these terms from the very basic perspective of a single electron function and a single point-charge nucleus, while the previous described how to ascend from there to the complete atom. The attention now meanders towards practical questions that arose during the implementation into the program code.

2.5 Practical Questions

2.5.1 Rotation of Spherical Harmonics

The biggest issue with applying the integrals (equation 2.28) to multipoles revolves around their relative orientations. The atomic multipoles in the Hansen-Coppens model inhabit local rotation frames that depend on chemical surroundings. This makes comparison between similar atoms possible, and it allows the creation of a database of multipolar parameters that one can transfer to chemically similar atoms. However, an electrostatic energy calculation demands the multipolar populations expressed in either a common or an interatomic rotation frame.

The rotation takes place in two steps. The first step is to express multipolar populations of all atoms in a common rotation frame. Nucleus-electron calculations rely on evaluating spherical harmonics directly, and the function `dlm_function` from `math.c` pre-supposes this common frame. Then, before every electron-electron integration for a particular pair of atoms, multipole populations switch over to the interatomic frame. After the calculation, they return back to the common frame for the next calculation.

To calculate rotated coefficients, one must first know how to rotate, and what are the rotation angles.[35] This requires some definitions.

Let X, Y, Z be the canonical Cartesian axes in three-dimensional space, and u, v, w an orthogonal set of axes with an arbitrary orientation. The two systems have a common origin. Let u_X, u_Y, u_Z denote the Cartesian coordinates of the base vector $\mathbf{u}$ in the direction u of the original (X, Y, Z) system, and by analogy v_X, v_Y, v_Z for $\mathbf{v}$, and w_X, w_Y, w_Z for $\mathbf{w}$.

[35] Prokopi, 2021.

The goal of the rotation routine is to rotate the axes of the original u, v, w coordinate system to those of the target canonical X, Y, Z one. Many sequences exist to perform this reorientation. The sequence used here is the w-v-w rotation sequence.[36] The rotation happens in three steps:

1. the rotation around the w axis brings the v axis into the XY plane,
2. the rotation around the v axis renders the w and Z axes parallel,
3. the second rotation around the w axis (now $\parallel Z$) makes the u and v axes parallel to X and Y respectively.

The three steps correspond to the three Euler angles:[37]

$$\alpha = \arctan \frac{v_x w_y - v_y w_x}{u_x w_y - u_y w_x}, \tag{2.33a}$$

$$\beta = \arctan \frac{w_x^2 + w_y^2}{w_z^2}, \tag{2.33b}$$

$$\gamma = -\arctan \frac{-w_x}{w_y}. \tag{2.33c}$$

A tricky case happens when w and Z are already parallel (but may point in opposing directions), so both x and y are already in the XY plane (rigorously defined as $w_x = w_y = 0$). When this happens, the formulas above enter into undefined territory. Their replacements are:

$$\alpha = 0, \tag{2.33d}$$

$$\beta = \begin{cases} 0 & w_z > 0 \\ \pi & w_z < 0, \end{cases} \tag{2.33e}$$

$$\gamma = -\arctan \frac{u_y}{u_x} \tag{2.33f}$$

This describes the transformation in the direction from original to Cartesian axes ($u, v, w \rightarrow X, Y, Z$). To go from the Cartesian axes to an interatomic system ($X, Y, Z \rightarrow u, v, w$), one needs to execute the transformation in the opposite direction. This is a simple exercise, where one calculates the angles α, β and γ as described, and then uses the **o**pposite angles $\alpha_o = -\gamma$, $\beta_o = -\beta$ and $\gamma_o = -\alpha$ in the next step.

With the Euler angles defined, one can turn to rotation of multipolar coefficients. Rotation of dipole coefficients represents the simplest case. It reflects a simple rotation of a vector in a three-dimensional space. One multiplies the vector of original $P_{1m\pm}$ coefficients with a rotation matrix $\mathbf{R}$ to obtain the rotated $P'_{1m\pm}$ coefficients:[38]

[36] This sequence is commonly known as z-y-z, but a change in notation from x, y,z to u, v, w alleviates confusion that symbols like x_X introduce.

[37] The fractions in equation 2.33 can not be supplied directly to the arctan function because it needs the information on the denominator and numerator signs to determine the correct quadrant (and thus the correct angle). One should use the `atan2(y, x)` function (with `y` as the denominator and `x` as the numerator) to correctly evaluate these angles.

[38] The sign at the end of $P_{lm\pm}$ denotes the sign of m, to avoid confusion of $P_{l\pm m}$ and the cumbersome $P_{l,\pm m}$.

$$\begin{bmatrix} P'_{11-} \\ P'_{10} \\ P'_{11+} \end{bmatrix} = \mathbf{R} \begin{bmatrix} P_{11-} \\ P_{10} \\ P_{11+} \end{bmatrix} \tag{2.34}$$

The rotation matrix flows from Euler angles:

$$\mathbf{R} = \begin{bmatrix} \cos\alpha\cos\gamma - \sin\alpha\cos\beta\sin\gamma & \sin\beta\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma \\ \sin\alpha\sin\beta & \cos\beta & \cos\alpha\sin\beta \\ \cos\alpha\sin\gamma + \sin\alpha\cos\beta\cos\gamma & -\sin\beta\cos\gamma & \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma \end{bmatrix} \tag{2.35}$$

Rotation of higher-order multipoles ($l > 1$) ascends from a three-dimensional transformation to a $(2l + 1)$-dimensional one:

$$\begin{bmatrix} P'_{ll-} \\ \vdots \\ P'_{l0} \\ \vdots \\ P'_{ll+} \end{bmatrix} = \mathbf{R} \begin{bmatrix} P_{\bar{ll}} \\ \vdots \\ P_{l0} \\ \vdots \\ P_{ll} \end{bmatrix} \tag{2.36}$$

The rotation in higher dimensions rests on the rotation of spherical harmonic d_{lm} functions themselves into a similar set of $d_{lm'}$ functions. It requires a definition of further terms to simplify the final matrix formulation:

$$\begin{aligned} d^{(l)}_{m'm}(\beta) = & \sqrt{\frac{(l+m')!}{(l-m')!}\frac{(l+m)!}{(l-m)!}}(-1)^{m'-m}\times \\ & \sum_k (-1)^k \binom{l+m}{k}\binom{l-m}{l-m'-k}[\cos(\beta/2)]^{2l-m'+m-2k}[\sin(\beta/2)]^{2k-m+m'} \end{aligned} \tag{2.37a}$$

$$\max(0, m-m') \le k \le \min(l-m', l+m) \tag{2.37b}$$

$$d^{(l)}_{-m'm}(\beta) = (-1)^{-l-m'} d^{(l)}_{m'm}(\beta+\pi) \tag{2.37c}$$

$$\phi_+ = |m|\gamma + |m'|\alpha \tag{2.37d}$$

$$\phi_- = |m|\gamma - |m'|\alpha \tag{2.37e}$$

The spherical harmonic $d_{lm}(\theta, \varphi)$ to $d_{lm'}(\theta', \varphi')$ rotation matrix is best defined by blocks based on the values of m and m':

$$\mathbf{M} = \begin{array}{c|ccc} & -l \le m \le -1 & m = 0 & 1 \le m \le l \\ -l \le m' & (-1)^{|m|+|m'|} d^{(l)}_{|m'||m|}(\beta)\cos(\phi_+) & [(-1)^{|m'|} d^{(l)}_{|m'|0}(\beta) + d^{(l)}_{-|m'|0}(\beta)] & (-1)^{m+|m'|} d^{(l)}_{|m'|m}(\beta)\sin(\phi_+) \\ m' \le -1 & -(-1)^{|m|} d^{(l)}_{-|m'||m|}(\beta)\cos(\phi_-) & \times \sin(|m'|\alpha)/\sqrt{2} & -(-1)^{m} d^{(l)}_{-|m'|m}(\beta)\sin(\phi_-) \\ \hline m' = 0 & (-1)^{|m|+1} d^{(l)}_{0|m|}(\beta)\sin(|m|\gamma)\sqrt{2} & d^{(l)}_{00}(\beta) & (-1)^{m} d^{(l)}_{0m}(\beta)\cos(m\gamma)\sqrt{2} \\ \hline 1 \le m' & (-1)^{|m|+m'+1} d^{(l)}_{m'|m|}(\beta)\sin(\phi_+) & [(-1)^{m'} d^{(l)}_{m'0}(\beta) + d^{(l)}_{\bar{m}'0}(\beta)] & (-1)^{m+m'} d^{(l)}_{m'm}(\beta)\cos(\phi_+) \\ m' \le l & +(-1)^{|m|+1} d^{(l)}_{-m'|m|}(\beta)\sin(\phi_-) & \times \cos(m'\alpha)/\sqrt{2} & +(-1)^{m} d^{(l)}_{-m'm}(\beta)\cos(\phi_-) \end{array} \tag{2.38}$$

Multipolar populations P_{lm} are linear coefficients multiplying spherical harmonics d_{lm}. It follows from linear algebra that a transformation applied on spherical harmonics d_{lm} is equivalent to an inverse-transpose transformation applied to linear coefficients P_{lm}. The rotation matrix for equation 2.36 comes out of this transformation:

$$\mathbf{R} = (\mathbf{M}^{-1})^{\top} \tag{2.39}$$

Rotation of dipoles d_{1m} has a particular property of orthogonality: $\mathbf{M} = (\mathbf{M}^{-1})^{\top}$. It can serve to verify calculation integrity. This does not apply to higher-order spherical harmonics.

2.5.2 Integral evaluation

Applying the calculations as described so far to heavier atoms (typically chlorine and iron, but also others) has revealed a weakness in using the main formula (equation 2.28) as given. The term $\epsilon_{\mathrm{B}} = e^{\zeta_{\mathrm{B}}\kappa_{\mathrm{B}}d}$ multiplies the integral from $a = d$ to ∞ in the relevant sub-equations. This term spills over the brim of a 64-bit floating point number in some cases (when ζ_{B} is large, and as the distance d between the two atom increases).

To solve this problem, one must revise the evaluation of a fundamental integral:

$$\int_a^{\infty} r^n e^{-br}\mathrm{d}r = \begin{cases} \frac{e^{-ab}}{b^{n+1}} \sum_{i=0}^{n} \frac{n!}{(n-i)!}(ab)^{n-i} & n \ge 0 \\ \text{set to } 0 & n = -1 \\ -(-b)^{-n-1} \sum_{i=1}^{-n-1} \frac{e^{-ab}(-n-i-1)!}{(ab)^{-n-i}(-n-1)!} & n \le -2 \end{cases} \tag{2.40}$$

The term e^{-ab} (marked in blue) differs in the three sub-equations for I_1, I_2 and I_3. It contains the antidote:

$$e^{\zeta_{\mathrm{B}}\kappa_{\mathrm{B}}d} \int_a^{\infty} r^n e^{-br}\mathrm{d}r = \begin{cases} \frac{e^{\zeta_{\mathrm{B}}\kappa_{\mathrm{B}}d-ab}}{b^{n+1}} \sum_{i=0}^{n} \frac{n!}{(n-i)!}(ab)^{n-i} & n \ge 0 \\ \text{set to } 0 & n = -1 \\ -(-b)^{-n-1} \sum_{i=1}^{-n-1} \frac{e^{\zeta_{\mathrm{B}}\kappa_{\mathrm{B}}d-ab}(-n-i-1)!}{(ab)^{-n-i}(-n-1)!} & n \le -2 \end{cases} \tag{2.41}$$

In the case of I_1 (equation 2.28b), $-ab = -\zeta_B\kappa_B d$, so the term becomes $e^{\zeta_B\kappa_B d - ab} = 1$. When evaluating I_2 or I_3 (equations 2.28c and 2.28d), $-ab = -\zeta_A\kappa_A - \zeta_B\kappa_B d$, so the term becomes $e^{\zeta_B\kappa_B d - ab} = e^{-\zeta_A\kappa_A d}$, which does not blow up.

The exponent $\zeta_B\kappa_B d - ab$ must therefore be evaluated before exponentiation, as on the right side of equation 2.41. This solves the problem. It also requires a different strategy to compute the integral from 0 to a (equation 2.30c), since the "standard" equation for that integral includes the integral from a to ∞.

2.6 Long Distance Approximation

2.6.1 The Approximation

The calculation at long distances does not have to be as heavy as when the atoms are close together because electron densities do not overlap in a considerable way. The Hansen-Coppens multipolar model can then give way to a simpler version of an atom.

In it, the nucleus and the spherical part of the electron cloud coalesce into a point charge. For a completely neutral atom, this charge is 0; if it has some ionic character ($P_{val} \neq N_{val}$ and/or $P_{00} \neq 0$ in terms of the Hansen-Coppens model), a small charge has to be considered ($q = P_{val} - N_{val} + P_{00}$). The interaction energy of two point **c**harges q_A and q_B from two atoms A and B at a distance d follows the well-known Coulomb law:[39]

$$E_{cc} = \frac{q_A q_B}{d} \tag{2.42}$$

This represents a major contribution at long distance.

The rest of the multipolar models, namely multipoles, coalesce into multipole moments at long distances. Buckingham proposed a long distance approximation for electrostatic energy based on multipolar moments.[40] A full implementation of the Buckingham multipolar moments (MM) would require converting multipole terms P_{lm} from the Hansen-Coppens model to the appropriate terms of traceless tensors (one for the dipolar, quadrupolar, octupolar *etc.* moments), and then perform tensor multiplications for all pairs of tensors (and include charge-tensor terms) of the two atoms A and B.[41]

The implementation of such a calculation was deemed time-consuming. Abandoning the guiding idea to coalesce multipoles into multipolar moments, an alternative emerged where the multipoles play a more straightforward role. Firstly, this approach sees Buckingham cross terms involving the point charge and multipole moments as point **c**harges in a potential field of **m**ultipoles:

$$E_{cm} = q_A \, \Phi_B(d) \, d_{lm}(\theta_A, \varphi_A) \tag{2.43}$$

[39] Coulomb, 1785, p. 529.
[40] Buckingham, 1959.
[41] The seventh chapter of Coppens, 1997, pp. 143–164 brings a broad overview of the full Buckingham (MM) calculation.

This term is analogous to the nucleus-electron term (equation 2.19), which permits the re-cycling of code used for that purpose. Like that term, it is also asymmetrical, so it has to be calculated a second time with indices inverted. The calculation of electrostatic potential $\Phi(r)d_{lm}(\theta_A, \varphi_A)$ is sufficiently fast and precise that it can substitute the equivalent Buckingham (tensor product) terms.

Since this approach does not involve coalescing multipoles into multipolar moments, the final **m**ultipole-**m**ultipole E_{mm} term remains the same as in the calculation at short distances (equation 2.28). The dipole-dipole term typically has a large impact on the overall energy, and other low-order multipole terms tend to be considerable. As a consequence, this term could profit from the speed-up by a complete implementation of the Buckingham approximation.

This approach carries the name *pseudo-Buckingham approximation* since it is similar to the Buckingham approximation in spirit. It is evaluated as:

$$E_{\text{pseudo-Buckingham}} = E_{cc} + E_{cm} + E_{mm} \tag{2.44}$$

2.6.2 Approximation Domain

One must ascertain the domain of applicability of the pseudo-Buckingham approximation. Figure 2.4 illustrates the split into three different distance domains: short, middle and long, fruitful for the upcoming discussion.

At short range, the pseudo-Buckingham approximation can not be valid due to overlap of electron clouds. Atoms as far away as 4–5 Å no longer have significant density overlap, so the approximate results agree well with the full integration. At long distances, the E_{mm} term (and the E_{ee} term in general) show signs of precision problems.

Two solutions exist for the low-end threshold d_{low} (figure 2.4) between low and middle distance. The first is a one-size-fits-all threshold:

$$d_{low} = d_0, \tag{2.45}$$

which should be set to at least 4 Å for faster calculations, or to 5 Å for a conservative estimate. The alternative is a sliding threshold based on van der Waals radii:

$$d_{low} = r_{vdW,A} + r_{vdW,B} + d_1, \tag{2.46}$$

with the additional margin d_1 of at least 1 Å for faster evaluations, or 2 Å for conservative estimates. This second threshold type can improve the speed of calculations for structures with many hydrogen atoms, as $2r_{vdW,H} \approx 2.4$ Å and $r_{vdW,H} + r_{vdW,B} \approx 2.8$ Å, while for most other atom pairs $r_{vdW,A} + r_{vdW,B} \approx 3.1$ to 3.6 Å.

At high distances, test calculations with high multipolar populations $P_{lm} = 0.3$ demonstrate at what point the calculations become numerically unstable, but also small enough to have no practical

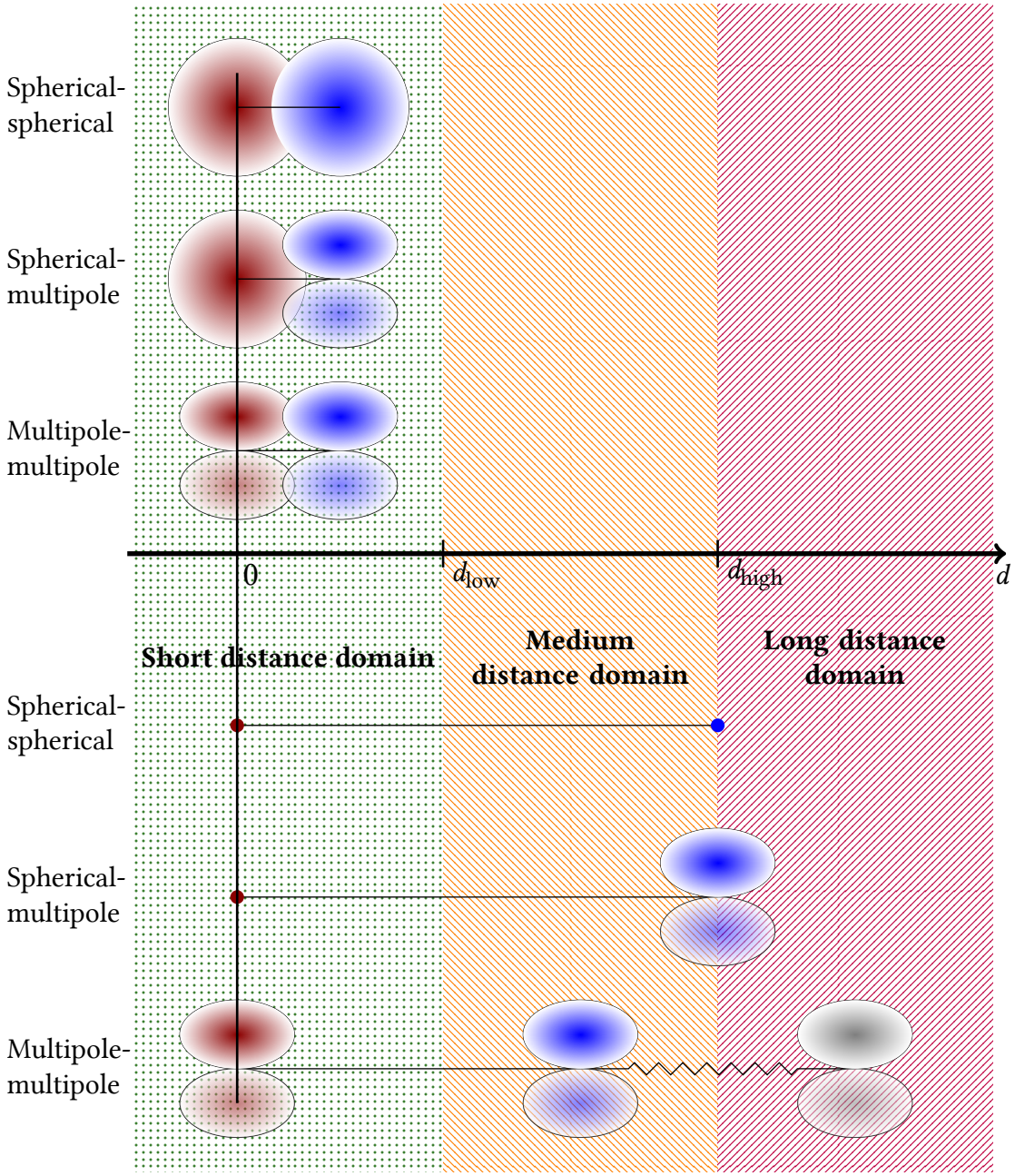


Figure 2.4. Different calculations in different domains. The red density is fixed at distance $d = 0$. Above the line: full integration of two densities typically happens when they are in the low distance domain due to significant density overlap. Below the line: a Buckingham-type approximation becomes valid at medium and long distance. Spherical densities become point charges. The spherical-spherical part reduces to two point charges, the spherical-multipolar part to a point charge in a multipolar potential field, but the multipolar-multipolar part remains the same as at short distance (full integration). However, multipolar-multipolar contributions at long distance are negligible and usually ignored (greyed-out dipole).

relevance. For all multipole-multipole pairs up to the hexadecapole level, the calculated energy drops well below $\pm 5 \cdot 10^{-3}$ kcal/mol long before numerical instabilities.

Table A.1 details a progressive set of cut-off values d_{high} applicable to different multipole pairs that guarantee fast execution with an acceptable maximum error of below $\pm 5 \cdot 10^{-3}$ kcal/mol. Dipole-dipole interactions which can have impactful contributions even at distances longer than 40 Å, but that distance (and the $\pm 5 \cdot 10^{-3}$ kcal/mol maximum error criterion) was a compromise to reduce computational time. The dipole-dipole cut-off boundary can be pushed further away without fear of numerical instability.

Estimated energies drop much faster for other multipole pairs (dipole-quadrupole, dipole-octupole, quadrupole-quadrupole *etc.*). The balance between computation time reduction and precision did not play such a big role when deciding cut-off distances for them. As a consequence, distances chosen for multipole pairs other than dipole-dipole typically entertain maximum errors much smaller than $\pm 5 \cdot 10^{-3}$ kcal/mol, making calculations with them even more precise.

This chapter presented the mathematics behind the analytical exact potential and pseudo-Buckingham approaches. The following reports on its manifestation in program code: the forging of *Charger*.

3 Charger

3.1 Résumé

Le chapitre précédent a introduit la partie mathématique qui soutient cette thèse. Les calculs invoqués par cette méthodologie sont répétitifs. Pour les déléguer à l'ordinateur, plus infatigable et plus fiable quand il faut effectuer des calculs répétés, une nouvelle librairie, `Charger`, a été conçue. Ce chapitre aborde les détails sur cette librairie.

Le chapitre commence sur les détails de la conception et de l'implémentation dans le code, expliquant certains choix particuliers provenant de la stratégie de développement du logiciel. Ensuite, le logiciel `MoProViewer`, très utile dans le cadre de la visualisation de molécules, a encadré `Charger` dans une interface dédiée pour effectuer les calculs de manière facile à utiliser.

`Charger` a été testé sur plusieurs systèmes moléculaires : d'abord sur des petites molécules ne contenant que carbone, hydrogène, oxygène et azote, puis sur des complexes protéine-ligand. Les deux points intéressants de ces tests sont la fiabilité de `Charger` (est-ce qu'il donne le même résultat comme les méthodes et logiciels établis ?) et sa rapidité (ce qui était le but principal derrière le projet). Le chapitre finit sur les problèmes et les limitations de `Charger` actuellement connus.

3.2 Introduction

The mathematical equations discussed in the previous chapter take form within a computer program called `Charger`. `Charger` is two things at once: a program library written in the programming language `C`, and the tool to calculate electrostatic interaction energies based on that library implemented in the `MoProViewer` program.

Before going into the details of the chapter, a short introduction to the `MoProSuite` is in order. Molecular `Pro`perties, or `MoPro`, is the cornerstone of the suite. It refines parameters of a Hansen-Coppens model based on structure factors from X-ray diffraction experiments. `VMoPro` is a multifunctional command-line tool to analyse the `MoPro`-generated Hansen-Coppens model of a crystalline solid. It notably includes a numerical implementation of the electrostatic energy calculation. Finally, `MoProViewer` is a molecular visualisation program that allows more intuitive atom selection and visualisation of analyisis results (*e.g.* 3D plots of electron density coloured by electrostatic potential). `Charger` was born into this family.

The first section of this chapter brings details on library contents (functions, modules *etc.*) to provide a path for integration into other projects. As the library is written in `C`, it packages data into `structs`. A `struct` is a complex data container that assembles disparate parts into a meaningful form.

The most important `struct` defined in `Charger` is `Atom`, a `struct` to represent the parameters of a Hansen-Coppens pseudo-atom.

Neither the `Atom` `struct` nor the `Charger` library in general assume any particular units in the calculation. They assume that nuclear charges Z, valence populations P_{val} and multipolar populations P_{lm} are given as multiples of elementary charge. If one gives atomic coordinates and ζ coefficients in atomic units (a_0 and a_0^{-1} respectively[1]), one obtains the calculated energy also in atomic units.[2]

The code library description serves mainly to link the physical contents discussed in the previous chapter with their equivalents in the program code. The next section examines the binding with `MoProViewer`, a molecular visualisation and multipolar model exploration program. The use of `MoProViewer` enables a faster and more intuitive input of a Hansen-Coppens model to a `Charger` calculation. Both front-end and back-end are examined.

Afterwards, an exploration of performance in terms of accuracy and calculation speed on simple and complex examples is in order. This mainly relies on a paper about a protein-ligand complex and some small-molecule benchmark systems, introduced briefly for this purpose. The protein-ligand complex will be revisited in the *Applications to Chemical and Biological Systems* part. Finally, the last section brings up some known unresolved issues with `Charger`.

3.3 Code Library

3.3.1 `Rust` and its Consequences

`Charger` was first incarnated as a `Rust` library. `Rust` is a programming language designed to force the programmer to write better code by preventing frequent mistakes that lead to memory failure or corruption, and thus to crashes. These guarding rails come at a twin price of frustration (`Rust` considering one's code unsafe, and one not knowing how to appease the compiler) and performance (slower execution time compared to the same code written in `C`).

The promise of reliability prompted the use of `Rust` as a base language for `Charger`. Once a sufficiently advanced and stable program in `Rust` formed, the `C` variant was built to improve performance and to facilitate integration in other projects, chiefly `MoProViewer` written in `C++`.

This choice influenced the design of the `C` version of `Charger`. A detour is thus necessary to explain some particular features of `Rust` that reflect in the design of the `C` library.

The first is a peculiar data type: `uintptr_t`. This is an integer type of the same size as the pointer on the chosen architecture (typically 32-bit or 64-bit, and the compiler adapts its size to the architecture choice).

[1] The symbol a_0 represents the Bohr radius, $a_0 = 5.291\,772\,109\,03(80) \times 10^{-11}$ m.

[2] The atomic unit of energy is hartree, $E_\text{h} = 4.359\,744\,722\,207\,1(85) \times 10^{-18}$ J $= 27.211\,386\,245\,988(53)$ eV. One must multiply the final result in E_h with 2625.499 639 479(88) or 627.509 474 063 0(69) to convert to kJ/mol or kcal/mol, respectively. The conversion comes from multiplying E_h with Avogadro's number, $N_\text{A} \equiv 6.022\,140\,76 \times 10^{23}$ mol^{-1}.

Listing 3.1. The two `Arc` `structs` used in the `C` version of `Charger`, found in `atom.h`. The contents of the two `structs` used within these `structs`, `Slater` and `Orbital`, are found in listing 3.3.

```
struct Arc_Slater {
  char padding[2*sizeof(uintptr_t)];
  struct Slater data;
};
struct Arc_Orbital {
  char padding[2*sizeof(uintptr_t)];
  struct Orbital data;
};
```

The second particular feature is the dynamically-sized container type, called a *vector*. It takes up three pieces of memory: a pointer `vec` to the first element, a counter `vec_cap` delimiting how much space belongs to `vec`, and a counter `vec_len` determining how many elements are present.[3]

The sequence `vec`, `vec_len`, `vec_cap` appears often, with a concrete name instead of `vec`. For example, the `multipoles` *vector* consists of a `multipole` pointer, a `multipole_len` number and a `multipole_cap` number.

The final feature is a double-pointer to an `Arc`, a `Rust` type intended for parallel computation. Some contents of an `Arc` are not truly relevant to `C`: the `Arc` `struct` contains padding bits to mirror the content from `Rust` without genuinely representing them. The `data` variable within an `Arc` contains the relevant underlying content. There are two types of `Arc` in the `C` library represented by two `structs`: an `Arc_Orbital` and an `Arc_Slater` (listing 3.1).

One should keep in mind this preliminary information as one explores the depths of the `Charger` library written in `C`. `Charger`'s purpose are chemical computations, so it has to have some notion of an atom.

3.3.2 The `Atom` `struct`

`Charger`'s foundational form seeks to delineate the inner workings of an atom. The fundamental `struct` within `Charger` is the `Atom`, defined in listing 3.2. The upcoming partitions reflect the split of an `Atom` into its constituent parts.

[3]The values within `vec_len` and `vec_cap` should be set to the length of `vec`, as `for`-loops within `Charger` go from 0 to `vec_len`. One should consider it safe to pass a pointer and the length of what this pointer points at to `Charger` without fear of memory trouble as `Charger` does not modify any input information.

Listing 3.2. The definition of the `Atom` `struct`.

```
struct Atom {
  char id[32];
  char element[32];
  char type_[32];
  double nuclear_charge;
  struct RadiusVector coordinates;
  struct Arc_Orbital **density;
  uintptr_t density_cap;
  uintptr_t density_len;
  double kappa;
  double pval_over_nval;
  double *kappa_prime;
  uintptr_t kappa_prime_cap;
  uintptr_t kappa_prime_len;
  struct Arc_Slater **multipoles;
  uintptr_t multipoles_cap;
  uintptr_t multipoles_len;
  double *plm;
  uintptr_t plm_cap;
  uintptr_t plm_len;
  double nval;
  double charge;
  double q_induced_over_nval;
  double *plm_induced;
  uintptr_t plm_induced_cap;
  uintptr_t plm_induced_len;
};
```

Labels

The `Atom` `struct` starts with three `char` arrays of 32-bit:[4]

- **`id`** a name used in output, mainly for debug purposes,
- **`element`** unused, but with possible future use,
- **`type_`** atom type, based on which atom properties are assigned (typically the element symbol).[5]

[4] 32-bit arrays are simpler to pass from one program to another through the foreign function interface, compared to dynamically-sized strings.

[5] The underscore (_) serves to distinguish the name of the variable from the `Rust` keyword `type`.

Listing 3.3. Auxiliary `structs` for the `Atom` `struct`. `RadiusVector` inhabits the header `math.h`, while other `structs` stack up next to `Atom` in `atom.h`.

```
struct RadiusVector {
  double x;
  double y;
  double z;
};
struct Orbital {
  char element_type[32];
  char label[32];
  struct Slater *slaters;
  uintptr_t slaters_cap;
  uintptr_t slaters_len;
  double population;
  bool valence;
};
struct Slater {
  double coefficient;
  double zeta;
  int n;
};
```

Basic Information

With the labels out of the way, the first pieces of information included in calculations are:

- **`nuclear_charge`** the nuclear charge Z of the specific atom, represented as a real number to avoid type conversion problems, and
- **`coordinates`** a `RadiusVector`, a `struct` containing the x, y, z coordinates of an atom (listing 3.3).

Orbital Density

The spherical orbital density representation resides within:

- **`density`** a *vector* of `Arc_Orbital` `structs`, which reflects the orbital structure of an atom.

An `Arc_Orbital` (listing 3.1) contains the data to represent one orbital. Each `Orbital` (listing 3.3) contains two 32-bit `char` arrays:

- **`element_type`** to match the `type_` of `Atom` – this is how `Charger` recognises which orbital density belongs to which atom, and
- **`label`** a simple orbital label (typically *1s*, *2s*) to simplify debugging.

Furthermore, it contains:

- **`slaters`** a *vector* of radial functions,

- **population** a real number representing the orbital population of a neutral orbital ($N_{i,\text{core}}$ or $N_{i,\text{val}}$, equations 2.4 and 2.5),
- **valence** a boolean designating the orbital as core (`false`) or valence (`true`).

A radial function (equation 2.10) takes the form of a `Slater struct` (listing 3.3), which contains the values of the three necessary elements:

- **coefficient** the linear coefficient $N_\text{r}N_\text{c}$,
- **zeta** the exponential coefficient ζ, and
- **n** the radial exponent n.

The linear combination of these radial functions (equation 2.7) makes up the square of the wave function (equation 2.9).

The information provided so far depicts an atom with a nucleus and neutral spherical density separated into core and valence orbitals.

The `Atom struct` (listing 3.2) further follows the Hansen-Coppens model (reproduced here from the previous chapter, with model parameters in bold font):

$$\rho_\text{atom}(\mathbf{r}) = P_\text{core}\rho_\text{core}(r) + \underline{P_\text{val}}\rho_\text{valence}(\underline{\kappa}r) + \sum_{l=0}^{l_\text{max}} R_l\left(\underline{\kappa'_l}r\right)\sum_{m=-l}^{l}\underline{P_{lm}}d_{lm}(\theta,\phi). \tag{2.3}$$

The `Atom struct` introduces κ (equation 2.3) and P_val (equations 2.3 and 2.6) as:

- **kappa** a real number, placeholder for κ, and
- **pval_over_nval** a real number to represent $P_\text{val}/N_\text{val}$, a direct multiplier to orbital density.

Multipole Density

The remaining part of the multipolar model are the multipoles, which follow in the `Atom struct` (listing 3.2) with:

- **kappa_prime** a *vector* of real numbers to represent the κ' value, or κ' values as a function of multipole level l (equation 2.3),[6]
- **multipoles** a *vector* of l `Arc_Slater` pointers (listing 3.1), one for each multipole level, to represent the radial part of multipole density (equation 2.3), and
- **plm** a *vector* of real numbers to hold multipole coefficients.

An `Arc_Slater struct` contains a radial function in the underlying `Slater struct`. This `struct` contains the same coefficients ($N_\text{r}N_\text{c}$, κ, ζ) as its spherical counterpart.

One can play with `multipoles_len` and `plm_len` to speed up calculations when entire multipole levels equal zero – for instance, if hydrogen atoms have coefficients up to hexadecapoles, but only their dipoles are non-zero. Care must be taken that `plm_len` always equals l^2, since `Charger` always expects the input of entire multipole levels, necessary for proper multipole rotation.

[6] If only one κ' value is used across all multipolar levels, `kappa_prime[0]` should be set to it, and `kappa_prime_len` = `kappa_prime_cap` = 1.

Long-Distance Approximation

Long-distance approximation requires two further values in the `Atom struct` (listing 3.2):

- **`nval`** a real number representing N_{val}, and
- **`charge`** a real number containing any additional charge that must be assigned to the atom type.

This charge does not refer to $N_{\text{val}} - P_{\text{val}} + P_{00}$, but rather to the charge of ionic atom types (*e.g.* Na^+ or I^-). If not for this number, `Charger` would not treat their formal charge correctly at long distance.[7] The calculation in the short-distance domain does not require this information since it relies directly on the orbital populations.

Induced Charge and Multipoles

The final elements in the `Atom struct` (listing 3.2) exist mainly for future compatibility in terms of induced charge and multipole moments. They represent:

- **`q_induced_over_nval`** the induced charge as $q_{\text{induced}}/N_{\text{val}}$, and
- **`plm_induiced`** a *vector* of induced multipole populations.

These values are integrated at all levels of calculation and only await non-zero input to be put to use.

3.3.3 Common Function Arguments

The `Atom struct` incorporates the mathematical portrayal of an atom, `Charger`'s animating spirit flows from its functions. They have many arguments in common, discussed here.

All top-level functions take one or two `Atoms` (`atom`, or `atom_a` and `atom_b`) as inputs. They further take several parameters called `constituents` (with suffixes `_a` and `_b` as appropriate). These can be:

- **a boolean** that represents whether or not to consider the nucleus,
- **4 booleans** that answers whether to use core, neutral valence, charged valence and induced charge valence densities,[8]
- **7 booleans** that determine which multipole levels to use,
- **7 booleans** with the suffix `_induced` that regulate which induced multipol levels to employ in the calculation.

Two more boolean arguments appear often at the atom level.

- **`long_distance`** argument controls whether to consider P_{00} multipoles or not. At long distance, the spherical P_{00} contributes to the point-charge term, and does not need to be integrated.
- **`x128bit`** argument, if set to `true`, switches the calculation to a 128-bit precision when the radial exponent $n \geq 6$.

[7] Taking Na^+ as an example, $N_{\text{val}} = 6$ because its 2p electrons are considered valence. When doing full integration, there are no 3s electron functions, and thus its charge is taken care of explicitly. However, at long distance, the whole spherical part of an atom is considered as a point with $N_{\text{val}} - P_{\text{val}} + P_{00}$ charge. This is correct for all neutral atoms, but Na^+ has an extra +1 charge that needs to be added to get the correct result.

[8] If either charged valence or induced charge valence is set to `true`, neutral valence is also considered as if set to `true`.

Listing 3.4. The `structs` (from `c_matrix.h`) and a function (from `c_matrix.c`) relevant for *C*-matrix storage. Donald Knuth suggested the number 2654435761.

```
#define CHARGER_CMATRIX_LENGTH 1117
struct Element_CMatrix {
  double* cmatrix;
  uintptr_t cmatrix_cap;
  uintptr_t cmatrix_len;
  int key;
};
struct Vec_Element_CMatrix {
  struct Element_CMatrix* element;
  uintptr_t element_cap;
  uintptr_t element_len;
};
struct HashTable_CMatrices {
  struct Vec_Element_CMatrix* table;
  uintptr_t table_cap;
  uintptr_t table_len;
  int length;
};
int hash_function(const int key, const int length) {
  return (int) (((long long) key * 2654435761)%(long long) length);
}
```

C-Matrices

The electron-electron terms require handling of *C*-matrices (equation 2.27, listing 3.4). Every $C_{l'}^{nlm}$-matrix has four unique identifiers, the quantum numbers n, l, m and l', which together form a *key*. The *key* has the form of an 8-digit number: the first pair of digits corresponds to n (01, 02, 03...), the second pair to l, the third pair to m, and the fourth pair to l'.

For easy access with dense storage, *C*-matrices fit into a *hash table*. An allegory helps to explain this concept. Imagine twenty-six book-shelves labelled with the twenty-six letters of the English alphabet. Each shelf contains the books from authors whose family names start with that letter. If one wishes to look for a book from Jules Verne (a *key*), he looks at the book-shelf "V" (the *hash* calculated from the *key*), and then examines each book until he finds the one from Mr. Verne. He may have to look through a dozen of books, but certainly a much smaller number than if all bookshelves contained books from any author with any family name.

The *C*-matrices lie on 1117 different "shelves", a number empirically observed to have a good compromise between many empty "shelves" (wasted space), and many packed "shelves" (wasted time when searching for the particular *C*-matrix).[9] The `Vec_Element_CMatrix` represents a "shelf" in program code, and `Element_CMatrix` represents an individual "book" (an individual *C*-matrix itself).

[9] The value `length` in `HashTable_CMatrices` must be set to `CHARGER_CMATRIX_LENGTH` (1117 by default) for the *hash table* to work correctly.

When one needs the $C_{l'}^{nlm}$-matrix, he first asks the function `index_to_int` to take the four quantum numbers and return the *key*. Then, the `hash_function` takes the *key* and turns it into a *hash*. This *hash* is the index of the `table` pointer in the `HashTable_CMatrices struct`, pointing to the "shelf" (the `Vec_Element_CMatrix`) where the desired *C*-matrix hides. This action is equivalent to going immediately to the book-shelf "V" to look for Mr. Verne's books.

Finally, one must look through the "books" (`Element_CMatrix`) until the `key` corresponds to the desired *key*: there, the `cmatrix` pointer contains a row-major representation of the sought-after *C*-matrix. *C*-matrices contain only integers, but some of them are too large for a 64-bit integer and all of them are used to multiply real numbers. Hence, the use of the real number (`double`) `cmatrix` pointer.

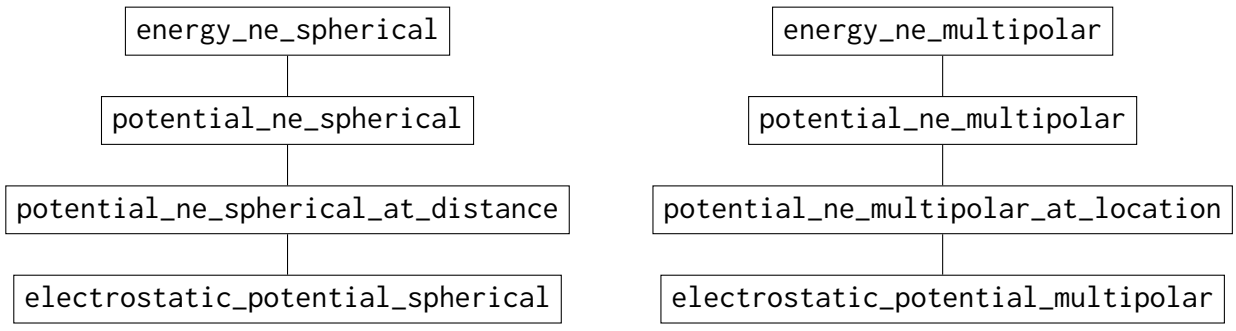
If a function asks for a pointer to a `double` called `cmatrix`, it looks for this pointer. The dimension of a $C_{l_A}^{n'_B l_B m_B}(i,j)$-matrix used in the principal expression (equation 2.28) is $n_B + l_B - m_B + l_A + 1$ by $n_B + l_A + 1$. The length of `cmatrix` does not need to be provided because the parameters n_B, l_B, m_B and l_A needed to determine its length are provided through other arguments.

With all these common parameters out of the way, the focus turns towards individual functions and their particular arguments. It forks in the nucleus-electron and electron-electron directions.

3.3.4 Nucleus-Electron Functions

The code file `jones1993.c` contains the functions to calculate the nucleus-electron term and electrostatic potential. It has four functions in a four-level hierarchy (figure 3.1):

- **`energy_ne_spherical`** calculates the spherical nucleus-electron energy by multiplying the result of `potential_ne_spherical` for atom A with the nuclear charge of atom B.
- **`potential_ne_spherical`** in turn calculates the distance between the atoms A and B provided to it, and calls the next function on the line,
- **`potential_ne_spherical_at_distance`** really handles the inner workings of the atom A; it can be called directly with a distance to calculate the electrostatic potential around an atom,



(a) The spherical hierarchy. (b) The multipolar hierarchy.

Figure 3.1. The parallel stacks of functions to compute the nucleus-electron term of the electrostatic interaction between two atoms.

- **`electrostatic_potential_spherical`** calculates the electrostatic potential generated by a spherically symmetrical Slater function with a specified κ value at a specified distance from the centre of the Slater function (equation 2.20 with $l = 0$).

The structure for the nucleus-multipolar electron functions mirrors the one for the nucleus-spherical electron term. These functions assume that the multipole populations inhabit the **canonical** Cartesian coordinate system.

The function on the third level in the nucleus-multipolar electron hierarchy, `potential_ne_multipolar_at_location`, asks not just for distance, but also for the θ and φ spherical angles (equation 2.13). The fourth level, `electrostatic_potential_multipolar`, needs the multipole level l specified, as well as `kappa`=κ', and all three spherical coordinates.

3.3.5 Electron-Electron Functions

The hierarchical organisation of electron-electron functions is different than that for the nucleus-electron functions. For starters, it has three layers:

- **the atom (top) layer** similar to the first three layers in the nucleus-electron hierarchy, whose functions take the `Atom structs` and separate them into pieces corresponding to single integrals,
- **the integral (middle) layer** corresponding to the last layer in the nucleus-electron hierarchy, whose functions evaluate the primary expression (equation 2.28),
- **the basic (bottom) layer** of basic integrals (Γ-functions), not needed in the nucleus-electron calculation.

The electron-electron hierarchy (figure 3.2) reflects the split of the electron-electron interaction problem into its spherical-spherical, spherical-multipolar and multipolar-multipolar sub-problems. This is already visible on the top level, with the functions `energy_ee_spherical`, `energy_ee_spherical_multipolar` and `energy_ee_multipolar`.

They take many common arguments: the *C*-matrix *hash table*, two atoms, their constituents (4- or 7-bit boolean arrays), and `long_distance` when necessary. They cut the atom up into individual

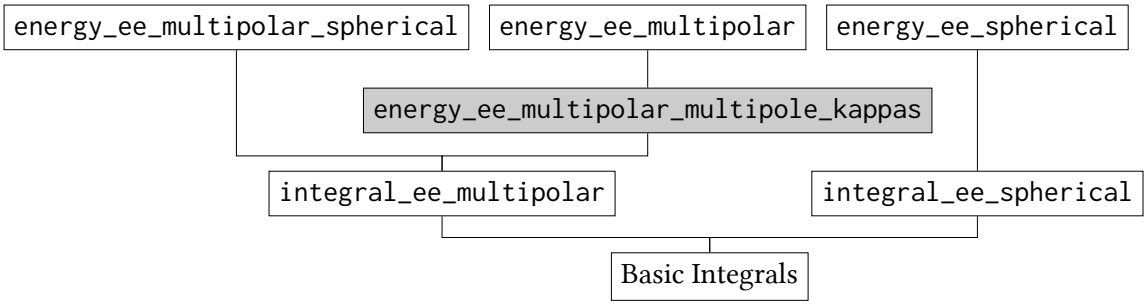


Figure 3.2. The stack of functions to compute the electron-electron term of the electrostatic interaction between two atoms. It has three layers and a greyed-out sublayer (`energy_ee_multipolar_multipole_kappas`) which is only accessed when necessary.

Slater functions (spherical or multipolar) and call for the relevant second-level function. They assume that the multipole populations inhabit an **interatomic** coordinate system.

The argument `z_in_direction_of_first_atom` is particular to the function `energy_ee_spherical_multipolar`. The interatomic coordinate system assumes that the z axis points towards the atom B, and the equations in the previous chapter rest on this assumption. However, the z axis can also point towards the atom A, since one has to switch the order of atoms A and B for the inverse calculation. To avoid re-calculating the interatomic system and all multipole coefficients, one can simply pass this argument as `true` in this case.

There is an intermediate layer between `energy_ee_multipolar` and the second hierarchy layer. It treats the special case when κ' is a function of the multipole level l for at least one atom in the atomic pair under momentary consideration. This happens rarely and increases computational time by about 23%, so a dedicated function `energy_ee_multipolar_multiple_kappas` takes over only when necessary. This function should never be called as a top-level function.

The second layer consists of two simple functions for treating the spherical and multipolar case: the `integral_ee_spherical` and the `integral_ee_multipolar` function. The first is similar to the `electrostatic_potential_spherical` function (final level of the nucleus-electron hierarchy) in that it takes Slater functions and κ values for two atoms. In addition, it takes only the relevant C-matrix (provided by its hierarchical superior). It then evaluates the equation 2.28 for the spherical case ($l_\mathrm{A} = l_\mathrm{B} = m_\mathrm{A} = m_\mathrm{B} = 0$) with the help of basic integral functions that evaluate the equation 2.30.

The `integral_ee_multipolar` function serves for both the spherical-multipolar and the multipolar-multipolar case. It asks for the same parameters as the `integral_ee_spherical` function (but only the distance, not spherical angles, since spherical harmonics do not appear in equation 2.28), and the l_A, l_B and `m_`=$|m|$ values, and the pre-computed values of the relevant basic integrals along with the largest calculated l_A value (`l_a_max`).

The pre-computed integrals improve the performance of the multipolar-multipolar calculation because multipoles use a single-ζ approach over different multipole levels l. This breaks down if κ' is a function of l, an already mentioned function treats that case. The integrals are pre-computed from $-2l_{\mathrm{A,\,max}}$ to $n_\mathrm{A} + n_\mathrm{B} + 3$, so `l_a_max` serves to identify the number of pre-computed integrals (n_A and n_B are given through `slater_a` and `slater_b`).

Basic Integral Functions

The functions to compute the Γ-function values (known here as basic integrals, equation 2.30), form the final layer of the electron-electron hierarchy. They merit their own title as they have other useful applications beyond `Charger`. Their code resides within `gamma_integrals.c`.

The first set of basic functions are the one-off functions `integrate_0_to_a`, `integrate_a_to_infinity` and `integrate_0_to_infinity`. They take the limit a (when necessary), the radial exponent n and the exponential factor b as the arguments. They evaluate the integrals $\iota(n) = \int_l^u r^n e^{-br}\,\mathrm{d}r$ within

the given lower (l) and upper (u) bounds by using the formulas within equation 2.30 and return the resulting value.

The use of these functions hampers performance. Firstly, values $\iota(n)$ of some integrals for fixed parameters a and b can be evaluated inductively, where $\iota(n+1) = f(n+1)\cdot\iota(n)$ for positive n, and $\iota(n-1) = f(n-1)\cdot\iota(n)$ for negative n.[10] Secondly, the intermediate-level functions (`integral_ee_spherical` and `integral_ee_multipolar`) need to use several times the values for $n = -2l_A, \dots, -1, 0, 1, \dots, n_{\max}$ and a pair of fixed parameters a and b.

Consequently, it pays off to create functions which calculate, for fixed a and b, a series of values $\iota(-2l_A), \dots, \iota(-1), \iota(0), \iota(1), \dots, \iota(n_{\max})$. The functions with the `_range` suffix in their name achieve this task. They take a pointer `output`, and they pre-suppose that it has enough reserved memory space to fit the requested range of results.

The `integrate_0_to_a` and `integrate_0_to_a_range` functions switch automatically to the more precise mode, following the criterion in the equation 2.30f. In this mode, they use the `lower_gamma_cfrac` function to evaluate the "otherwise" case in equation 2.30c, and `m_cfrac` using the modified Lentz's method to evaluate the continued fraction from the equation 2.30e.[11] They do so on individual basis, with no margin to improve performance.

3.4 MoProViewer BINDING

`MoProViewer` is the flagship program from `MoProSuite`, the set of refinement and molecule exploration tools developed at the CRM² laboratory. `MoProViewer` displays molecules and charge density properties (critical points, basins, *etc.*) and performs various calculations directly or by calling supplementary programs such as `VMoPro`.

`Charger` relies heavily on the selection options from `MoProViewer` (figure 3.3). These involve the possibility to select a group of atoms, retain this selection saved, and select another group of atoms. `Charger` then uses these two groups (saved and current selection) as inputs.

3.4.1 Front End

The minimalistic `Charger` interface in `MoProViewer` facilitates electrostatic energy calculations. It is located on the *Energy* window (figure 3.4a). It contains the options to select constituents of atoms from saved and current selection (nucleus, core electrons, valence neutral/charged, P_{00}, dipoles, induced dipoles, quadrupoles, octupoles, hexadecapoles, 32-poles and 64-poles[12]). These options apply to all atoms in each selection. This allows one to calculate, as an example, the interaction energy of the total protein density and the dipole ligand density.

[10] The function f is a placeholder for the iterative element that depends on the boundaries of the integral ι. Irrespective of those boundaries, it is always possible to find such a function by looking at the formulas in equation 2.30.

[11] Winkler, 1993, pp. 169–171,Abergel and Moisan, 2020.

[12] Calculations with 32-poles and 64-poles have not been tested.

The calculation starts by clicking the *Run Charger* button. Once it ends, the four boxes display the resulting electrostatic interaction energy in atomic units (e^2/a_0) and "crystallographic energy units" ($e^2/Å$), as well as the molar energy (the same energy divided by Avogadro's number) in kcal/mol and kJ/mol.

The log file saves these energies and molar energies. It complements them with identifying information (atom names), the constituents of both groups, group size, number of atoms in the short and long distance domains, and calculation time.

The `Charger Protein` tab (figure 3.4b) provides a simple interface adapted for specific protein-ligand calculations. It demands a saved selection containing ligand atoms, and a choice of first and last residue number. Clicking *Run Charger* starts a calculation in which `MoProViewer`:

1. silently, behind the scenes, selects all atoms of the first chosen residue,
2. starts a `Charger` calculation between the residue and the ligand,
3. reports the answer in the log file,
4. de-selects the residue and selects the next one,
5. repeats steps 2.-4. until the last chosen residue.

If a residue of a given number is missing, or atoms from that residue overlap with the saved selection, `MoProViewer` skips the calculation. While the driving force behind making this module were the described protein-ligand calculations, it can find applications outside that domain.

Finally, the `Charger Properties` tab (figure 3.4c) gives access to common modifiers of `Charger` calculations:

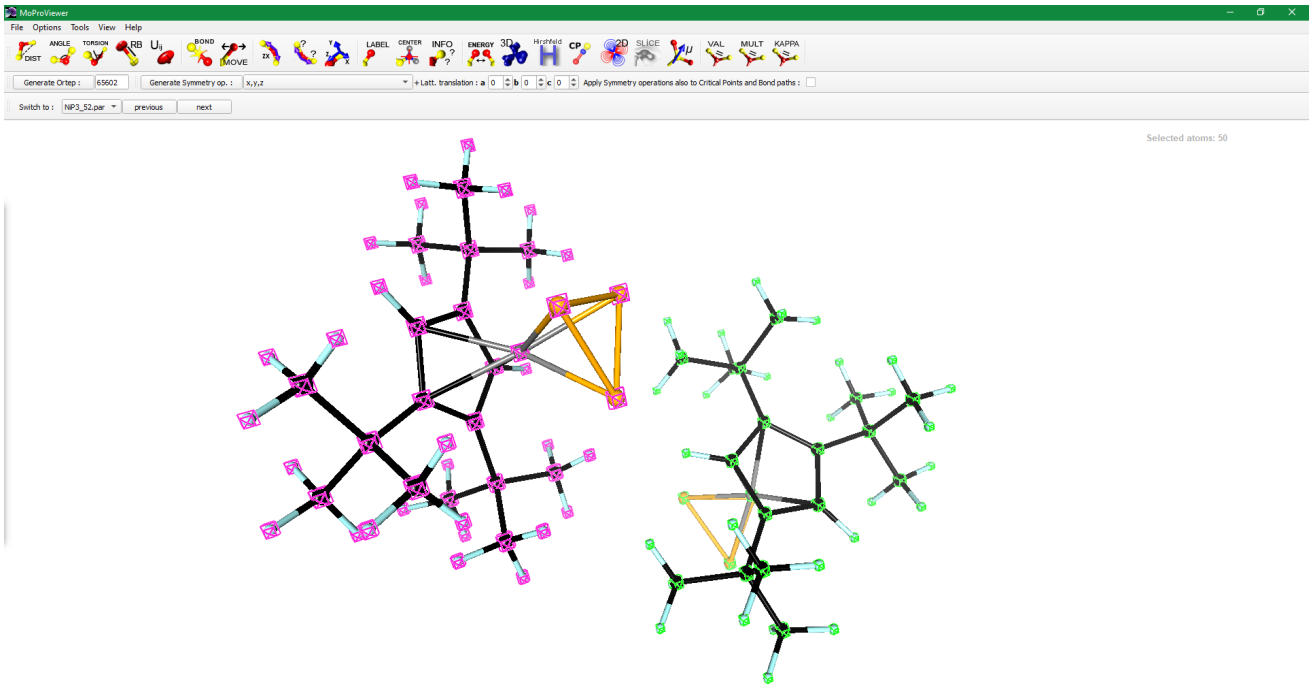
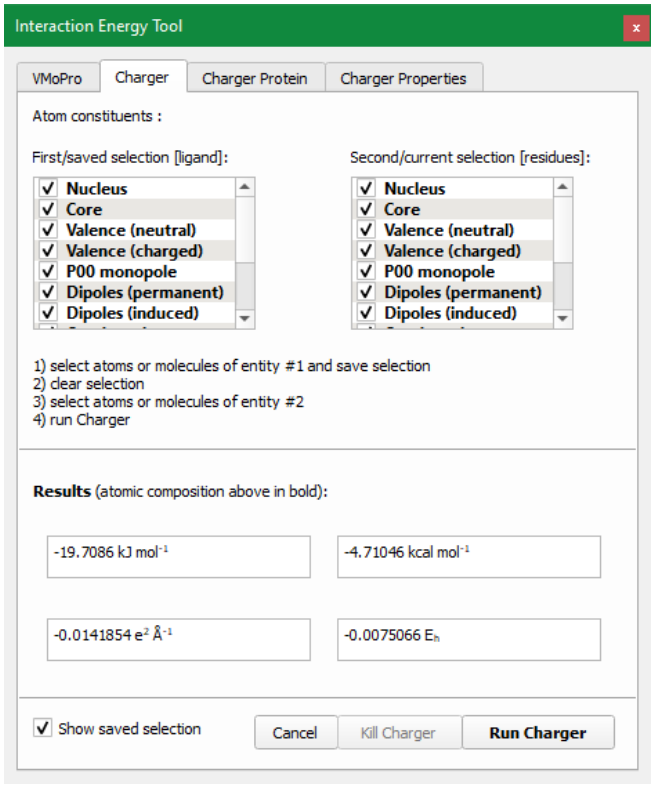
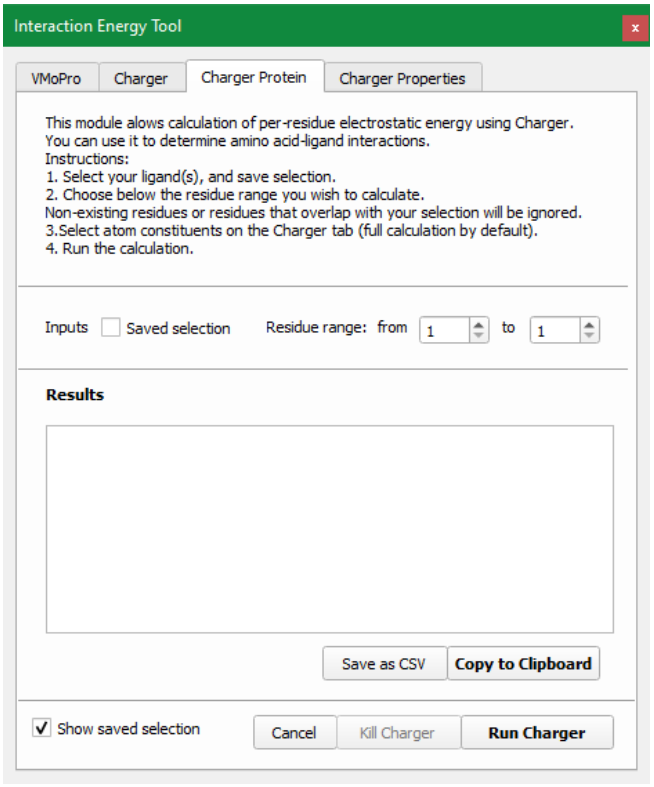


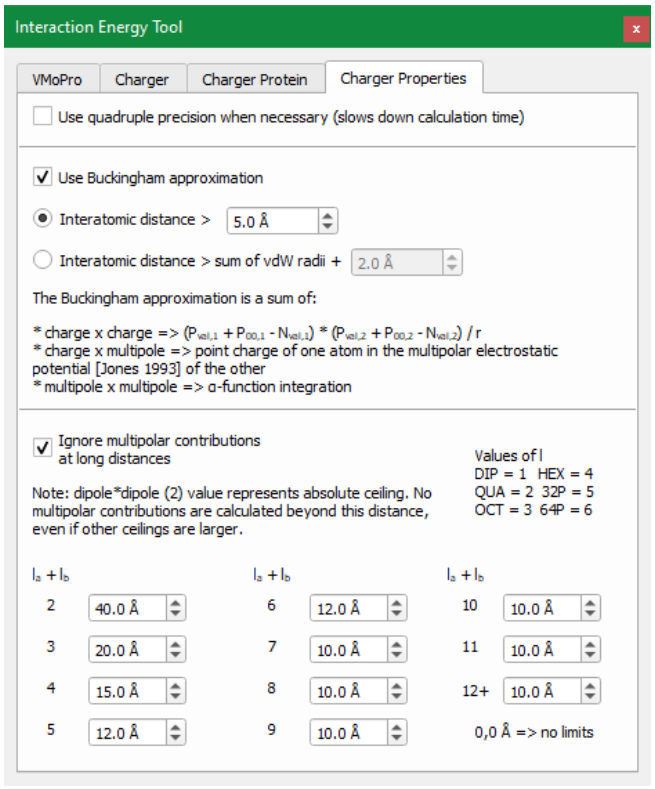
Figure 3.3. The `MoProViewer` window with two $Cp'''NiP_3$ molecules selected for a `Charger` calculation (purple: current selection, green: saved selection). Figure 3.4a shows the result of this calculation.



(a)



(b)



(c)

Figure 3.4. The `Charger` dialogs within `MoProViewer`: (a) `Charger` proper, with the interaction energy between two $Cp'''NiP_3$ complexes as shown on figure 3.3, (b) `Charger Protein`, (c) `Charger Properties` with default cut-off values as in table A.1.

- the option to use 128-bit (quadruple precision floating points) calculations for integrals with $n \geq 6$,
- the pseudo-Buckingham domain threshold, fixed or dependent on van der Waals radii of the atom pair (equations 2.45 and 2.46),
- the cut-off distances for multipole-multipole calculations:
  - as a function of the sum of multipolar levels $l_A + l_B$,
  - one can find the cut-off distances set as default in table A.1 (under the *precise* column),
  - $l_A + l_B = 12$ (a 64-pole × 64-pole calculation, the maximum level implemented by `MoProSuite` at the time of writing the thesis) applies automatically to $l_A + l_B > 12$ for future compatibility,
  - when the $l_A + l_B = 2$ threshold is passed, no further multipole-multipole calculations take place.

3.4.2 Back End

`MoProViewer` contains a layer of functions to re-arrange the data from existing `C++ classes` to an input suitable for `Charger` (figure 3.5). They also load unavailable supplementary information from data files. The information is loaded only once, and `boolean` switches ensure that the preparation step does not repeat.

The `Charger_Calculation` function rests at the top of the interface stack. This represents the abstracted calculation of electrostatic interaction energy for two selections with pre-loaded parameters. It calls the topmost functions in the `Charger` nucleus-electron and electron-electron stacks (described in subsections 3.3.4 and 3.3.5) for all pairs of atoms such that the first comes from the saved, and the second from the current selection.

The `Charger_Calculation` handles the switch from the short-distance to the long-distance regimen. It passes on long distance cut-offs to `Charger` functions. It does not call the multipole-multipole function when the atoms are too far off for any multipole-multipole calculations, thus saving considerable time.

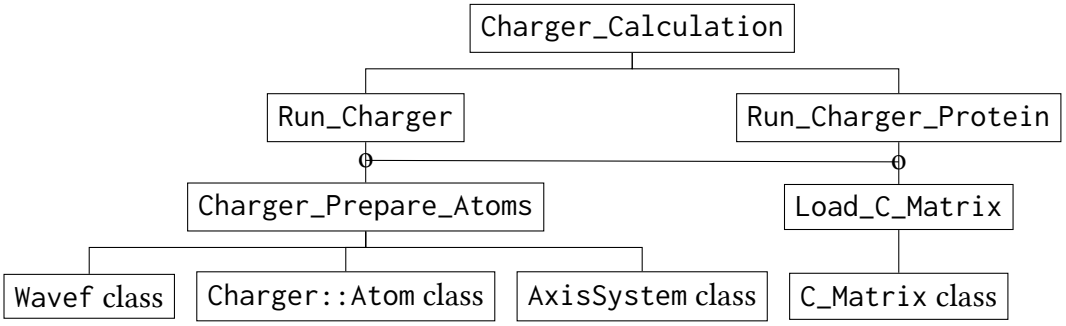


Figure 3.5. The stack of functions `MoProViewer` back end uses to prepare the input for `Charger`. The function at the top of the stack calls `Charger` library functions (figures 3.1 and 3.2).

Finally, Charger_Calculation handles switching between molecular and interatomic coordinate systems. Before calling the calculation of electron-electron terms containing multipoles, it uses the AxisSystem class to rotate multipole coefficients P_{lm} into the interatomic system for the current pair of atoms. Once these calculations finish, it restores the coefficients to the molecular coordinate system.

The front-end Run buttons on Charger and Charger Protein tabs call the functions with the same name. These functions first call the preparatory functions on the lower level, and pass the prepared atomic models to Charger_Calculation afterwards. They also take care of window and log-file updates.

Charger_Prepare_Atoms is the main and the most complex piece of the back-end machinery. It first charges an object of the Molecule class to load information from the WAVEF and MoPro.tab data-files though the Wavef class. The Properties window of the front end provides an option to specify the paths to these files.

The WAVEF data file contains tabulated wave function coefficients for spherical orbital density (equations 2.9 and 2.10).[13] These are converted into density coefficients (equation 2.7 – it is a square of a sum, so the number of individual radial terms rises from n to $n(n+1)/2$ for a particular orbital).

The MoPro.tab file contains information on which orbitals belong to the core, and which to the valence layer, as well as their neutral populations. It also carries the total neutral valence population of the chemically neutral species to derive the correct charge for long-distance calculations.[14]

The MoPro.tab file further contains information on multipolar density. The Molecule class had a facility to read these coefficients prior to building the Charger back end. It uses this facility and reformulates the coefficients in a Charger-friendly input format.

The loaded and correctly stored orbital and multipolar density model resides within the Wavef class. The procedure stores the largest radial exponent $n_{\max}$ it found in any (spherical or multipolar) radial density function. It loads from WAVEF and MoPro.tab files the coefficients for all atom types that exist in the molecular model. An atom type may exist in the molecular model, but not in either of the data files. If no selection contains that atom type, a Charger calculation proceeds normally.

The MoProViewer C++ Atom class contains a Load_Charger_Atom function which uses information contained within Atom to construct a charger::Atom struct (listing 3.2). The function assumes that all information from WAVEF and MoPro.tab files is available. Charger can use the created struct as input for its calculations.

[13] The atomic data tables list wave function $n_{\mathrm{wf}} = n + 1$ values.

[14] *E.g.* for Na^+, MoPro.tab lists its total neutral valence population as the total neutral valence population of Na. If one sums up all valence populations as stated in MoPro.tab for Na^+, the result is one fewer than its stated total valence population. This translates to a point charge of $+1 + N_{\mathrm{val}} - P_{\mathrm{val}} - P_{00}$, which represents the spherical part of Na^+ in long-distance calculations.

The Load_Charger_Atom function handles the multipolar populations P_{lm}. The Atom class sees them as multipoles with MoPro order[15]and sign,[16] given in terms of the local coordinate system defined by the scrutinized atom and two close neighbours. Charger needs an input in l, m-order[17] and with standard signs, and Load_Charger_Atom ensures that it gets them.

The multipole rotation procedure (equation 2.36) also assumes these multipoles are given in the l, m-order and standard signs.[18] Multipoles are first rotated into a Cartesian rotation frame and saved as such in the Atom class. This allows the Charger_Calculation function to effortlessly convert back from interatomic to molecular coordinate systems by swapping a *vector* pointer instead of performing matrix multiplication. The Euler angles and rotation matrix for rotation come from the AxisSystem class.

This concludes the preparation procedure for a charger::Atom struct. The final remaining piece for a Charger calculation, specifically for electron-electron integrals (equation 2.28), are the C-matrices. They reside within the c_matrix_n.mat files found in a single directory (n is the major quantum number related to C-matrices, equation 2.27). Just like data files, one can modify this path through the Properties window of the MoProViewer front end.

The Load_C_Matrix function first figures out how many C-matrices a Charger calculation might need. The Wavef class saved the n_max integer, the largest radial exponent $n_{\max}$ found while loading the information for relevant atoms from the WAVEF and MoPro.tab data files. The matrices necessary for a calculation have the principal quantum number $n \leq n_{\max} + 3$. Load_C_Matrix then calls the extend function from the C_Matrix class, which handles everything to load C-matrices up to n for the molecule in question. C-matrices up to $n = 11$ suffice for most calculations.

3.5 PERFORMANCE

This section switches gears and touches on two major topics. The first is the issue of accuracy: do the results of the method described here conform to those of known older methods? Then comes speed: is the new method faster, what can one expect in realistic scenarios? The article *A rush to explore protein-ligand electrostatic interaction energy with* Charger (reproduced in the appendix, under chapter A) goes over these topics in detail. A short preface summarises key methodology points and chemical systems under investigation, and also defines particular terms often used in that article.

[15]The MoPro order of multipoles is P_{00}, P_{11+} (P_x), P_{11-} (P_y), P_{10} (P_z), P_{20}, P_{21+}, P_{21-}, P_{22+}, P_{22-}, P_{30}, P_{31+}, P_{31-}, P_{32+}, P_{32-}, P_{33+}, P_{33-}, P_{40}, P_{41+}, P_{41-}, P_{42+}, P_{42-}, P_{43+}, P_{43-}, P_{44+}, P_{44-}, ...

[16]MoPro base code contains peculiar definitions of the $d_{33-}(\theta, \varphi)$ and $d_{43-}(\theta, \varphi)$ spherical harmonics. Their sign is the opposite of the one found in common definitions of the same spherical harmonics.

[17]The order tolerated by Charger is first by increasing l, then by increasing m: P_{00}, P_{11-}, P_{10}, P_{11+}, P_{22-}, P_{21-}, P_{20}, P_{21+}, P_{22+}, ...

[18]One could change the order of rows and columns in the rotation matrix to make it applicable to any order of multipoles, even the MoPro one. Similarly, changing the signs of relevant rows and columns can make the rotation matrix applicable to multipoles with different signs. However, it is easier to re-order a one-dimensional vector and change the sign of two elements within it, than to change the order and sign of several rows and columns in the rotation matrix.

t

3.5.1 Preface

First in line are the investigated chemical systems. The article starts by considering benchmark systems proposed by Nguyen, Kisiel and Volkov.[19] These systems are homo- or heterodimers containing the following molecules: water, small alcohols or amines, amino-acids and polypeptides up to a dodecapeptide. Some polypeptides (*e.g.* Leu-enkephalin) exist in nature, while others contain amino acids not found in living beings. Successful calculations on these systems serve as a proof-of-concept for calculations on larger systems such as proteins.

All benchmark systems come from crystal structures. The smallest (a water-water dimer) contains 3 x 3 atom pairs, while the largest (a dodecapeptide dimer) contains 225 x 225 atom pairs. The multipole model parameters come from a databank transfer, as described in the methodology section of the paper (section A.3).

The article focuses mainly on a series of protein-ligand complexes: *Tv*GSTO3S with four different hydroxybenzophenones. *Tv*GSTO3S is a glutathione transferase, a group of proteins commonly abbreviated to GST. The chemical and biological overview of its interaction with hydroxybenzophenones has to wait until the next part, which brings forward several applications of electrostatic interaction energy calculations.

For the purpose of this section, one must know that GST is a homodimeric protein, and that each monomer in the dimer contains two active sites: one that binds a hydroxybenzophenone, and the other which binds glutathione (a glutamate-cysteine-glycine tripeptide).

One *Tv*GSTO3S monomer contains 240 amino acids. The performed calculations use amino acids 2–239, to avoid problems with modelling terminal amino acids. The N-terminal methionine 1 is cleaved in the biologically active protein, and the C-terminal glutamate 240 was a victim of a charge transfer bug resolved after the calculation. They are far away from the ligands and do not affect binding energies significantly. One protein monomer thus consists of 3706 atoms, and a dimer has 7412 atoms. Three benzophenone ligands contain 26, and the fourth 27 atoms, while the glutathione ligand contains 36 atoms. The methodology section of the paper (section A.3) gives the details of protein structure preparation for calculations, including multipole model parameter transfer.

With chemical systems out of the way, one can focus on different methods for calculating electrostatic interaction energies (equation 1.2). The Nguyen, Kisiel and Volkov article describes the methodology to calculate electrostatic energy, giving it the name *analytical **e**xact electrostatic potential/multipolar moments* (a**E**P/MM). Multipolar moments is another name for the long-distance Buckingham approximation. They implemented this method in an in-house version of `XDPROP`, a charge density analysis program from the `XD` suite. This `XDPROP` version is not publicly available.[20]

The previous chapter details a derivation of electrostatic interaction energy calculation method largely based on aEP/MM, and this chapter expands upon its implementation within the `Charger`

[19] Nguyen, Kisiel and Volkov, 2018.
[20] Ibid.

program code. The article in appendix A names it the *analytical exact electrostatic potential/pseudo-multipolar moments* method (aEP/pMM).

Finally, the article by Volkov, Koritsanszky and Coppens presents the original EP/MM method.[21] Due to its numerical approach to the same integral (equation 1.2), Nguyen, Kisiel and Volkov renamed it to *numerical exact electrostatic potential/multipolar moments* (nEP/MM). Implementations of this method exist in `VMoPro`[22] (a part of `MoPro` program suite) and `XDPROP`[23] (a part of `XD` program suite), both of which are publicly available. The article in appendix A uses the `VMoPro`-nEP/MM implementation as a reference method. For the sake of brevity and clarity, the remainder of this section will refer to the `XDPROP`-aEP/MM implementation as aEP/MM, the `Charger`-aEP/pMM as aEP/pMM, and the `VMoPro`-nEP/MM as nEP/MM.

Before concluding, a short note on cut-offs. The cut-offs described in the *Long Distance Approximation* section of the previous chapter (section 2.6, given in table A.1) provide a reasonable balance between calculation speed and precision, and are labelled the "precise" cut-offs.

Another set of cut-offs (also given in table A.1), which cuts all multipole-multipole calculations off beyond 15 Å, proved to obtain reliable results, while improving computation speed. It was baptised the "rapid" calculation.

3.5.2 Accuracy

The calculation accuracy is a highly relevant topic. One must consider two reference points: the values calculated by Nguyen, Kisiel and Volkov on small molecule benchmark systems,[24] as well as the values on the same benchmark systems and a large protein-ligand system, values coming from the established nEP/MM method.

The Bland-Altman plot provides a graphical summary of the comparison between two methods that ought to yield the same results. The x axis of this plot shows values of the studied quantity on the real scale (the scale of energies for the current purpose). In particular, it holds the arithmetic mean between the results of the two methods (for electrostatic interaction energy E calculations, this is E_{mean}). The y axis shows the difference between the tested and the reference method ($\Delta E = E_{\text{aEP/pMM}} - E_{\text{nEP/MM}}$ represents this quantity here).

The plot further shows the estimated mean of energy differences $m(\Delta E)$ and delimits the space between $m(\Delta E) - 2\sigma(\Delta E)$ and $m(\Delta E) + 2\sigma(\Delta E)$, where $\sigma(\Delta E)$ represents the standard deviation of energy differences. For comparison purposes, the graph that resulted after applying the tested aEP/pMM and reference nEP/MM methods (figure A.2) has two datasets: the blue (using precise cut-offs) and the red (using rapid cut-offs, table A.1).

The difference between energies obtained from tested aEP/pMM and reference nEP/MM calculations amounts to -0.02 ± 0.39 kcal/mol ($m \pm s$) if using precise cut-offs. Rapid cut-offs yield slightly

[21] Volkov, Koritsanszky and Coppens, 2004.
[22] Jelsch *et al.*, 2005; Fournier *et al.*, 2009.
[23] Volkov, Macchi *et al.*, 2016.
[24] Nguyen, Kisiel and Volkov, 2018.

less congruent results, as expected, with a mean difference of -0.06 ± 0.60 kcal/mol ($m \pm s$). These differences refer to electrostatic interaction energies of benchmark dimers (in squares) and GST-benzophenone complexes (with and without applied dipole polarisation – encircled and naked, figure A.2). This is a good sign that the different methods yield similar results.

The Bland-Altman plot of electrostatic interaction energies (figure A.2) presents clearly, however, that the energy difference depends on the value of calculated energy, and this dependence is roughly affine. The additive nature of the molecular model (the linear combinations from equation 2.1) implies the addition of energies on all levels of calculation (from the integral of two Slater functions, across the integral of two orbital densities, all the way up to the integral of two molecular densities). It also implies the addition of errors resulting from fundamental approximations inherent in the model (for instance, the α-function represented through the C-matrix), approximations when computing basic integrals (equation 2.30), and numerical errors coming from floating-point precision limits on any level of calculation.

If a 1% relative error were coming from the described sources at a low calculation level (for instance, an error when calculating basic integrals between two Slater radial functions), it would propagate upward to higher levels, resulting in a roughly 1% error in the final result (the integral of two molecular densities). This would make the energy difference $\Delta E \approx 0.5$ kcal/mol for $E \approx -50$ kcal/mol, $\Delta E \approx 1.0$ kcal/mol for $E \approx -100$ kcal/mol, and so on, which would explain the roughly affine dependency of ΔE on E_{mean}. One can estimate the actual relative error from the plot (figure A.2) to somewhat less than 1%.

A small relative error on a low calculation level has the biggest impact when the final result is close to zero. The reason comes from imperfect cancellation of intermediate results, which also suffer from the already listed errors. The result might have a wrong order of magnitude, or even the wrong sign. One should regard sceptically all estimates of energy below ±2 kJ/mol or ±0.5 kcal/mol. This is roughly the same order of magnitude as observed standard deviations ($\sigma(\Delta E)$).

Comparing to the benchmark, the differences are within 0.2 kcal/mol (tables A.2 and 3.1). They presumably come from numerical integration errors in the nEP/MM method. The energy values are also largely comparable to those available from Nguyen, Kisiel and Volkov (the aEP/MM method, figure A.1).[25] This is in spite of the fact that aEP/pMM calculations relied on electron density models with a different set of multipolar parameters than the aEP/MM method. In relative terms, the biggest differences come from dimers whose interaction energy value is lower than ±10 kcal/mol, expectedly from significantly different underlying models.

Another point of comparison were the `Charger` calculations with AMBER03 charges, commonly used in molecular dynamics simulations on biological macromolecules.[26] When they were used as spherical valence charges in the context of a Hansen-Coppens model, the results match well for almost all tested dimers (Tables A.2 and 3.2). One can find similarities between these results and the

[25] Nguyen, Kisiel and Volkov, 2018, tables 2 and 3, pp. 530, 531.
[26] Duan *et al.*, 2003.

Table 3.1. **Electrostatic interaction energies (in kcal·mol^{-1}) for interactions in benchmark systems used by Nguyen *et al.* (2018).**

Dimer designation	Monomer A	Monomer B	Charger precise calculation *	Charger rapid calculation *	VMoPro	NKV paper **	δ_{NKV} / % $[(E_{Charger\ p.} - E_{NKV})/E_{NKV}]$
Gly1	Glycine	Glycine	−27.9	−27.9	−27.8	−27.7	0.7
Gly2			−7.2	−7.2	−7.3	−7.0	2.9
Gly3			−19.8	−19.8	−19.8	−20.9	−5.3
Gly4			−35.7	−35.7	−35.9	−39.1	−8.7
Gly5			10.1	10.1	10.1	11.1	−9.0
Gly6			−5.5	−5.5	−5.6	−5.6	−1.8
Lenk1	Leu-enkephalin	L-enk	−64.4†	−64.4†	−64.5†	−74.8†	−†
Lenk2						−37.5†	
Lenk3			4.5	4.5	4.5	−3.8	−218
Lenk4	Leu-enkephalin	Water	−13.3	−13.3	−13.3	−15.3	−13.1
Lenk5			−14.4	−14.4	−14.4	−14.9	−3.4
Lenk6			−14.7	−14.7	−14.8	−10.7	37.4
Lenk7			−14.4	−14.4	−14.4	−12.9	11.6
Lenk8			−8.6	−8.6	−8.6	−8.5	1.9
Lenk9			−9.1	−9.1	−9.1	−5.7	59.7
Lenk10			−5.4	−5.4	−5.5	−7.1	−24.0
Lenk11	Water	Water	−10.6	−10.6	−10.5	−10.3	2.9
Non3††	Nonap.†††	DMA††	1.6	1.6	1.6	−9.6	−116
Dec1	Decap.†††	Decap.†††	−36.2	−36.2	−36.4	−47.4	−23.6
Dec2		Methanol	−15.6	−15.6	−15.6	−14.0	11.4
Dec3		Methanol	−8.8	−8.8	−8.8	−7.0	25.7
Dec4		Water	−10.4	−10.4	−10.6	−10.7	−2.8
Dec5		Water	−2.3	−2.3	−2.3	−3.0	−23.3
Dod1	Dodecap.†††	Dodecap.†††	−43.5	−43.1	−43.5	−48.4	−10.1
Dod2		2-butanone	−8.8	−8.8	−8.9	−3.5	151.4
Dod3		2-butanone	−1.7	−1.8	−1.8	−3.1	−45.2

* Precise and rapid cut-off criteria are reported in Table A.1.

** Refers to values as given in (Nguyen *et al.*, 2018) and converted from kJ·mol^{-1} to kcal·mol^{-1}.The values are not the same because the models are not completely equal, specifically the multipolar parameters.

*** Electrostatic interaction energy calculation by *Charger* based on an AMBER point-charge model. Excluded results had problems with model transfer.

† The file Lenk_dimer_1.xyz in the Supplementary Materials of (Nguyen *et al.*, 2018) does not contain a Leu-enkephalin dimer. Instead, it contains the same glycine dimer as the file Gly_dimer_6.xyz. We were therefore only able to obtain results for Lenk2, which are listed in the table. We do not give the relative differences, as we are not sure whether the file Lenk_dimer_2.xyz contains the dimer listed as Lenk1 or as Lenk2.

†† Nonapeptide – dimethylacetamide dimer. The files for the other two dimers (Non1 and Non2) contained three molecules, so they were not included.

††† Nonapeptide: Boc-(l-Leu-l-Leu-Aib)2-l-Leu-d-Leu-Aib-OMe; decapeptide: Boc-(l-Leu-Aib)5-OMe; dodecapeptide: Boc-l-Leu-l-Leu-Aib-(d-Leu-d-Leu-Aib)2-l-Leu-l-Leu-Aib-OMe; Boc = *tert*-butoxycarbonyl – (CH3)3**C**O**OC**–, Leu = leucine, Aib = 2-methylalanine. There are no parameters in the ELMAM2 database for the quaternary carbon and the carbonyl group in Boc (in bold), nor for the quaternary carbon C_α of 2-methylalanine.

We did not calculate a polarised electron density model for these calculations.

Table 3.2. **Electrostatic interaction energies (in kcal·mol^{-1}) for interactions in benchmark systems used by Nguyen *et al.* (2018).** Comparison with the AMBER point-charge model.

Dimer designation	Monomer A	Monomer B	Charger*	AMBER**	δ_{AMBER} / %	AMBER P_{val}**	$\delta_{\text{Amber } P_{\text{val}}}$ / %
Gly1	Glycine	Glycine	−27.9	−12.0	133	−23.2	20
Gly2			−7.2	−0.4	1700	−8.9	−19
Gly3			−19.8	−11.7	69	−16.7	19
Gly4			−35.7	−18.5	93	−30.7	16
Gly5			10.1	6.7	51	3.8	166
Gly6			−5.5	−2.3	139	−4.5	22
Lenk2†	Leu-enkephalin	L-enk	−64.4	−39.0	65	−51.9	24
Lenk3			4.5	2.7	67	−2.7	−267
Lenk4	Leu-enkephalin	Water††	−13.3	−8.8	51	−18.6	−28
Lenk5			−14.4	−10.4	38	−16.9	−15
Lenk6			−14.7	−7.1	107	−13.1	12
Lenk7			−14.4	−8.5	69	−13.7	5
Lenk8			−8.6	−3.6	139	−8.6	0
Lenk9			−9.1	−3.9	133	−11.6	−22
Lenk10			−5.4	−5.9	−8	−9.1	−41
Lenk11	Water††	Water††	−10.6	−7.5	41	−12.6	−16

* Precise calculation, as reported in table 3.1.

** Electrostatic interaction energy calculation by *Charger* based on an AMBER point-charge model. The AMBER column contains results for models where AMBER point charges are interpreted as point charges. The AMBER P_{val} column contains results for models in which AMBER point charges are converted to P_{val} values from the Hansen-Coppens model, and full spherical integration is performed for all atoms closer than 5 Å apart.

*** $\delta_{\text{AMBER}} = (E_{\text{Charger}} - -E_{\text{AMBER}})/E_{\text{AMBER}}$, $\delta_{\text{AMBER } P_{\text{val}}} = (E_{\text{AMBER}} - -E_{\text{AMBER } P_{\text{val}}})/E_{\text{AMBER } P_{\text{val}}}$.

† See note † from table 3.1 to explain why Lenk1 is missing.

†† Point charges of a TIP3P water molecule used.

Dimers excluded from this table (compared to Table 3.1) have parameters missing.

results of Kumar, Bojarowski *et al.* They performed an in-depth comparison of several charge density models, including AMBER point charges and Hansen-Coppens multipoles, on the S66 and JSCH-2005 molecular datasets.[27] The agreement is however poorer if one uses AMBER charges as point charges, which, due to the missing penetration energy, yield significantly lower energy values. This shows the value of `Charger` electrostatic calculations in the biomacromolecular context.

Finally, the article *An unusual interaction of a lone pair and an π-electrons in a quinoid dianion* (reproduced in appendix B) brings forward several `Charger`-aEP/pMM calculations (subsection B.3 of the article). The nEP/MM method implemented in `VMoPro` served to verify those results as well. The differences $E_{\text{nEP/MM}} - E_{\text{aEP/pMM}}$ are similar to those observed in the first article.

This article studied the DMSO···DHQ^{2-} complex formed in the crystal phase. It brings the direct comparison between the `Charger` electrostatic interaction energy and the total electronic energy from a DFT calculation. The first, between DHQ^{2-} and DMSO, is −19.1 kJ/mol, while the second, describing the formation of the DMSO···DHQ^{2-} complex, is more than double, −42.8 kJ/mol. This

[27] Kumar, Bojarowski *et al.*, 2014.

should warn against the use of Charger energies as total binding energies, especially when electrostatic interactions make a small contribution to overall binding energy.

3.5.3 Speed

Charger was conceived as a tool to calculate electrostatic interaction energy faster than the available ones. Exploring calculation speed thus makes an important part of evaluating Charger as a useful tool for exploring molecular contacts and interactions.

A direct comparison between calculation times of Charger's aEP/pMM and XDPROP's aEP/MM method are not possible at present because the XDPROP version containing aEP/MM is not available to the public. An indirect comparison, using values published in the Nguyen, Kisiel and Volkov paper, remains possible, with the serious limiting factor of not using the same processor for both calculations.[28]

Table 3.3 gives raw values of computational times for dimers explored by Nguyen, Kisiel and Volkov side by side with those from Charger calculations. Charger-aEP/pMM computation times have at most the same order of magnitude as XDPROP-aEP/MM. Charger appears relatively faster with smaller molecules which have more nearby atoms. The performance of Charger and XDPROP is further compared on figure 3.6, showing how well they scale up as a function of atom pairs involved in the calculation.

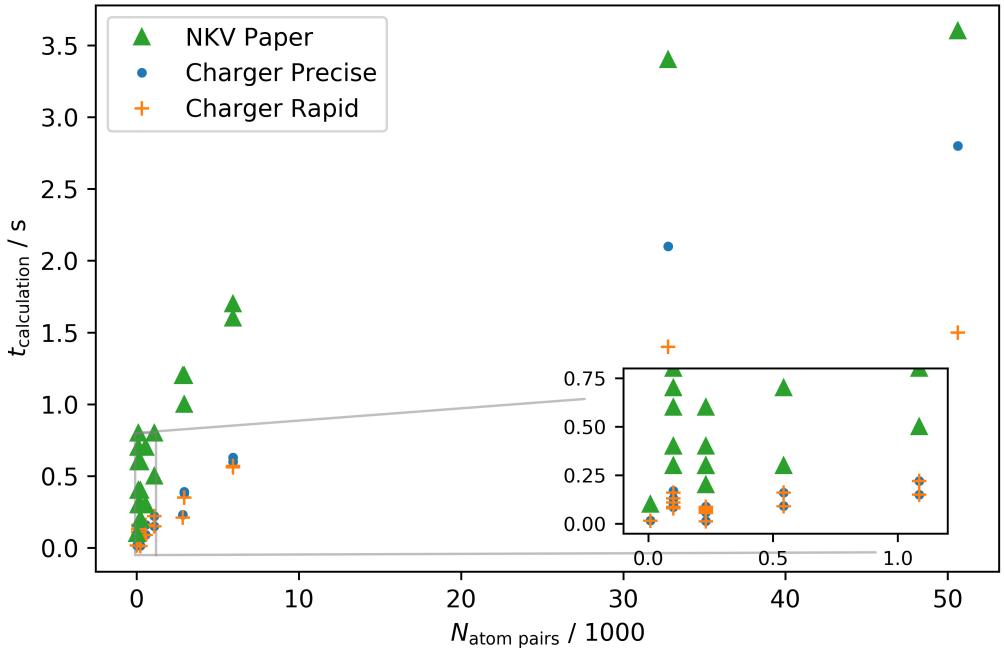


Figure 3.6. The calculation time as a function of number of atom pairs in benchmark systems, showing algorithm complexity as a function of input size. Inset: zoom-in on the region near the origin.

[28] Nguyen, Kisiel and Volkov, 2018, table 4, p. 534.

Table 3.3. Calculation times (in seconds) for electrostatic interaction energies in benchmark systems.

Dimer designation	Monomer A	Monomer B	Charger precise calculation*	Charger rapid calculation*	NKV paper**
Gly1	Glycine	Glycine	0.085	0.080	0.3
Gly2			0.094	0.087	0.4
Gly3			0.13	0.11	0.6
Gly4			0.17	0.16	0.8
Gly5			0.13	0.13	0.7
Gly6			0.087	0.086	0.4
Lenk1	Leu-enkephalin	L-enk	–†	–†	3.3
Lenk2			0.60	0.56	1.6
Lenk3			0.63	0.57	1.7
Lenk4	L-enk	Water	0.077	0.070	0.3
Lenk5			0.058	0.056	0.2
Lenk6			0.078	0.076	0.3
Lenk7			0.069	0.068	0.3
Lenk8			0.068	0.067	0.3
Lenk9			0.014	0.013	0.6
Lenk10			0.089	0.087	0.4
Lenk11	Water	Water	0.017	0.016	0.1
Non3***	Nonapep.††	DMA***	0.23	0.21	1.2
Dec1	Decapep.††	Decapep.††	2.1	1.4	3.4
Dec2		Methanol	0.15	0.15	0.5
Dec3		Methanol	0.22	0.22	0.8
Dec4		Water	0.16	0.16	0.7
Dec5		Water	0.091	0.090	0.3
Dod1	Dodecapep.††	Dodecapep.††	2.8	1.5	3.6
Dod2		2-butanone	0.39	0.35	1.0
Dod3		2-butanone	0.38	0.35	1.2

The processor used for `Charger` and `VMoPro` calculations was Intel® Core™ i7-8700 @ 3.20 GHz, Passmark (processor benchmark): 13.106.

* Precise and rapid cut-off criteria are reported in Table A.1.

** Refers to values as given in Nguyen, Kisiel and Volkov, 2018 for an aEP/MM calculation. Nguyen, Kisiel and Volkov used AMD Opteron 6348 @ 2.80 GHz, Passmark: 9.151.

*** Nonapeptide – dimethylacetamide dimer. The files for the other two dimers (Non1 and Non2) contained three molecules, so they were not included.

† The file Lenk_dimer_1 in the Supplementary Materials of Nguyen, Kisiel and Volkov, 2018 contain a glycine dimer instead of a Leu-enkephalin dimer. The energy value reproduced here serves mainly for comparison with Lenk2.

†† Nonapeptide: Boc-(l-Leu-l-Leu-Aib)2-l-Leu-d-Leu-Aib-OMe; decapeptide: Boc-(l-Leu-Aib)5-OMe; dodecapeptide: Boc-l-Leu-l-Leu-Aib-(d-Leu-d-Leu-Aib)2-l-Leu-l-Leu-Aib-OMe; Boc = tert-butoxycarbonyl – $(CH_3)_3COOC$–, Leu = leucine, Aib = 2-methylalanine. There are no parameters in the ELMAM2 database for the quaternary carbon and the carbonyl group in Boc (in bold), nor for the quaternary carbon Cα of 2-methylalanine.

The `VMoPro` executable containing a numerical EP/MM implementation facilitated the comparison between its nEP/MM and analytical EP/pMM from `Charger`. The increase in computation speed (figure A.3) correlates heavily with the number of atom pairs in the low-distance domain (depicted on figure 2.4). Full integration is obligatory in that domain due to significant density overlap. The largest improvement comes from increasing the integration performance in that domain, by the substitution of a numerical cubature integration in nEP – a triple integral (equation 2.22) – with an analytical integral of radial functions in aEP (equation 2.28) – a single integral, due to orthonormality of spherical harmonics (equation 2.31).

This substitution increases performance from 4 to 150 times for the studied systems (benchmark, GST-hydroxybenzophenone and GST-glutathione dimers, figure A.3). The aEP/pMM calculation is about 4 times faster than the nEP/MM one for specific GST-glutathione dimers. The *Tv*GSTO3S crystal structures contain two GST proteins with two binding sites, in which hydroxybenzophenone and glutathione ligands can nest. The glutathione binding site of one GST rests near the other, so that the glutathione interacts somewhat with the other GST from the crystal structure. There are about 6 atom pairs in 10 000 in the low-distance domain (figure 2.4) of one such GST-glutathione dimer, but that is already enough to provoke a slight speed improvement.

On the other side, 150 times speed improvement comes from benchmark systems in which almost all, if not all, atoms fall into the low-distance domain. The figure A.3 hides cases with many atoms in the low-distance domain, since it is difficult to say whether their improvement (which can go up to 200 times) comes from their small size (*e.g.* the water-water dimer has 3x3 atoms) or the fact that they have so many atoms close together.

`Charger`-aEP/pMM is not always faster than the reference `VMoPro`-nEP/MM. The second method takes the crown when no atom pairs inhabit the low-distance domain. This provides a clear sign that multipolar moments (MM) works faster than pseudo-multipolar moments, opening a clear point for improvement. Despite this, a `Charger` calculation should still be preferred from within `MoProViewer` because it does not need to re-load the structure from file to start the calculation. `VMoPro` always reads the input file before a calculation, which usually takes longer than the time one gains by using the faster MM method.

3.6 Issues

Charger has several unresolved issues that limit the scope of its applicability. Some should be taken into account when using it, but most are relevant for future development.

As noted at the end of the previous section, Charger does not contain a full multipolar moments (MM) calculation at long distances. This decreases computation speed, moreso for large systems with many distant atoms. The proposed cut-offs (table A.1) mitigate this problem imperfectly.

A more serious danger at long distances comes from numerical instability of integral evaluation. It has been observed empirically that this instability comes from integrals where $n \geq 6$. The radial exponent n acts on the interatomic distance d, and the term d^{n_l} suffers significantly when both d and n are large. This affects multipole calculations with large multipole levels (the condition $n > l$ must always hold, and typically $n > l + 2$) and heavier atoms with many orbitals with high n.[29]

The quadruple-precision Charger implementation does not solve this issue. Cut-offs effectively prevent the problems for multipole-multipole calculations at far distances, but no solution currently exists for short-distance calculations involving spherical density. Atoms up to nickel are known to give correct results, but test calculations on compounds with bromine and iodine gave nonsensical results – order of magnitude in MJ/mol. It is difficult to estimate whether the atoms between Ni and Br would give correct, slightly off or visibly incorrect results, so one should verify them with appropriate VMoPro calculations whenever possible. One can however rely on the Charger results for small organic molecules and proteins.

Finally, MoProViewer crashes upon repeated Charger calculations from the MoProViewer interface. This problem appears related to multipole calculations with Charger, and seems to come from Charger itself, or possibly the interaction between MoProViewer and Charger. It happens either at repeated calculations (multiple clicks on either Run Charger button, figure 3.4) or during a single calculation (*e.g.* for two protein monomers).

This problem prevents a large-scale exploration of protein-protein interactions and inconveniences the user.[30] The calculations on small systems (which do not crash after every click on Run Charger) work normally after restarting MoProViewer. There is no sign that the bug affects the exactness of calculation, as they always yield the same result (repeated runs by repeated clicks on Run Charger), and this result is similar to the one obtained from the alternative (numerical) method to calculate the same integral (VMoPro-nEP/MM).

[29] One must square the wave function to get the density, and so the maximum n value for a spherical density is $2(n_{\mathrm{wf,max}-1)}$. The term $n_{\mathrm{wf,max}}$ represents the value listed in the atomic tables based on *in vacuo* calculations; refer to section 2.3 for details. The value $2(n_{\mathrm{wf,max}-1)}$ goes up to 8 for heavy atoms.

[30] One can calculate protein-protein interaction energies with some trickery: select one whole protein, and a part of the other, small enough that MoProViewer does not crash; after re-starting MoProViewer, select another part of the second protein, and so on. The total energy is the sum of obtained energies.

Part II

Applications to Chemical and Biological Systems

4 Introduction

The first part of this thesis deals with the theoretical overview of the analytical exact potential / pseudo-multipolar moments (aEP/pMM) method implemented within the `Charger` program library. It explores the mathematical background of the Hansen-Coppens model, the Löwdin α-function, the Jones electrostatic potential and their applications to obtain the analytical solution of the exact potential method. It also peruses the Buckingham's multipolar moments methodology to incarnate a similar method (pMM). It discusses how `Charger` integrates with `MoProViewer` and how well it performs in terms of both accuracy and calculation speed.

With the confidence in results of `Charger` calculations that this elaboration has so far provided, one can turn to possible applications. The field of applications is wide, as one can in principle explore any and all compounds for which a Hansen-Coppens model exists. `Charger` has problems with compounds containing heavier atoms due to high n and ζ coefficients in Slater radial functions: this still leaves a large variety of compounds (anything with elemental composition up to nickel at least), and some quite interesting ones at that.

The previous chapter already brought forward the first application of `Charger`: the glutathione transferase *Tv*GSTO3S with glutathione and different benzophenone ligands. These protein-ligand systems, for which experimental structures are available, exemplify a textbook example of predominantly electrostatic interactions. `Charger` helped to find correlations between electrostatic interaction energies and experimental values and to point out which amino acid residues contribute significantly to binding. Furthermore, `Charger` pointed out a residue that is not favourable to binding from an electrostatic perspective.

The second example is a curious compound $Li_2(DHQ)(DMSO)_2(H_2O)_2$ (LiDHQ for short), found by a serendipitous re-crystallisation of the dihydroxyquinonate (DHQ^{2-}) salt of lithium from dimethylsulfoxyde (DMSO). The particular feature of this compound is an unexpected short contact between the DMSO sulfur atom and the dihydroxyquinonate carbon atoms. To characterise this unusual interaction, `Charger` joined the array of usual tools to explore a Hansen-Coppens model from a topology perspective (critical points, basins, *etc.*). This was also an opportunity to compare `Charger` directly with DFT calculations of the total interaction energy, as mentioned previously.

The final example is of a new compound that *débuts* at the Hansen-Coppens stage: the $Cp'''NiP_3$, an organometallic complex consisting of a nickel atom between a 1,3,4-*tris*(*tert*-butyl)cyclopentadiene ring (Cp''') and a triphosphate ring (P_3). The $Cp'''NiP_3$ complex is explored with the same set of tools as LiDHQ.

A further destiny awaits this complex and a well-known compound – the chloranilic acid (H_2CA), joining the fray in the last moment. `Charger` calculates the electrostatic energy of crystal packing for both compounds. This is the newest way to utilise `Charger` to explore crystal structures.

Overall, this part brings forward several of the myriad possible applications of `Charger` to highlight its wide pertinence: a protein-ligand complex, an organic moiety, a salt of an organic anion co-crystallised with an organic solvent, and an organometallic complex. They should not be thought of as proofs that `Charger` works infallibly in all contexts, but rather as pointers to the broad world of possibilities before it.

5 Glutathione Transferase and Its Ligands

5.1 Résumé

L'analyse d'un système protéine-ligand était le premier défi réel pour `Charger`. Le modèle choisi a été une glutathion transferase (GST) qui présentait une affinité pour les benzophénones comme ligands. L'opportunité qu'offrait ce modèle était l'existence de données expérimentales, à savoir des données thermodynamiques en solution et quatre structures cristallographiques de complexes protéine-ligand.

L'analyse énergétique des complexes protéine-ligand, réalisée avec l'aide de `Charger`, est probante car les énergies d'interaction électrostatique sont corrélées avec les données thermodynamiques. Cela montre que `Charger` est plus qu'un outil théorique. De plus, les résultats révèlent les acides aminés qui stabilisent les ligands dans la poche du site actif, mais aussi les résidus qui gênent leur fixation d'un point de vue électrostatique.

`Charger` offre donc un nouveau moyen pour analyser l'interaction entre deux partenaires et même l'améliorer en proposant des modifications. Dans le cadre de cette étude, la modélisation d'une mutation ponctuelle semble améliorer la fixation du ligand en réduisant un obstacle électrostatique qui n'avait pas été détecté dans l'analyse cristallographique. Dans cette approche particulière on cherche à éliminer un mauvais contact plutôt qu'à enrichir le nombre et/ou la qualité des contacts favorables. Une analyse détaillée du champ électrostatique soutient la notion d'obstacle diminué.

5.2 Introduction

The protein-ligand system under examination has already been briefly introduced in a previous chapter on accuracy and speed of `Charger` calculations. That chapter viewed the system as a collection of atoms divided into two groups, and the discussion pivoted around the adequacy of `Charger` for the calculation. Here the focus turns to what one can learn about the protein-ligand system from a biochemical and biological perspective using `Charger`.

The protein in the system is a glutathione transferase (GST[1]). GSTs serve an important purpose in the cellular detoxification system by catalysing the reversible binding of glutathione to electrophilic substrates. They also break up cellular toxins by oxydising them. The Ω class of fungal GSTs degrade lignins, a class of natural polyphenyl compounds that solidify wood bark.[2]

[1] The old name for proteins belonging to this family is glutathione S-transferases.
[2] Sheehan *et al.*, 2001.

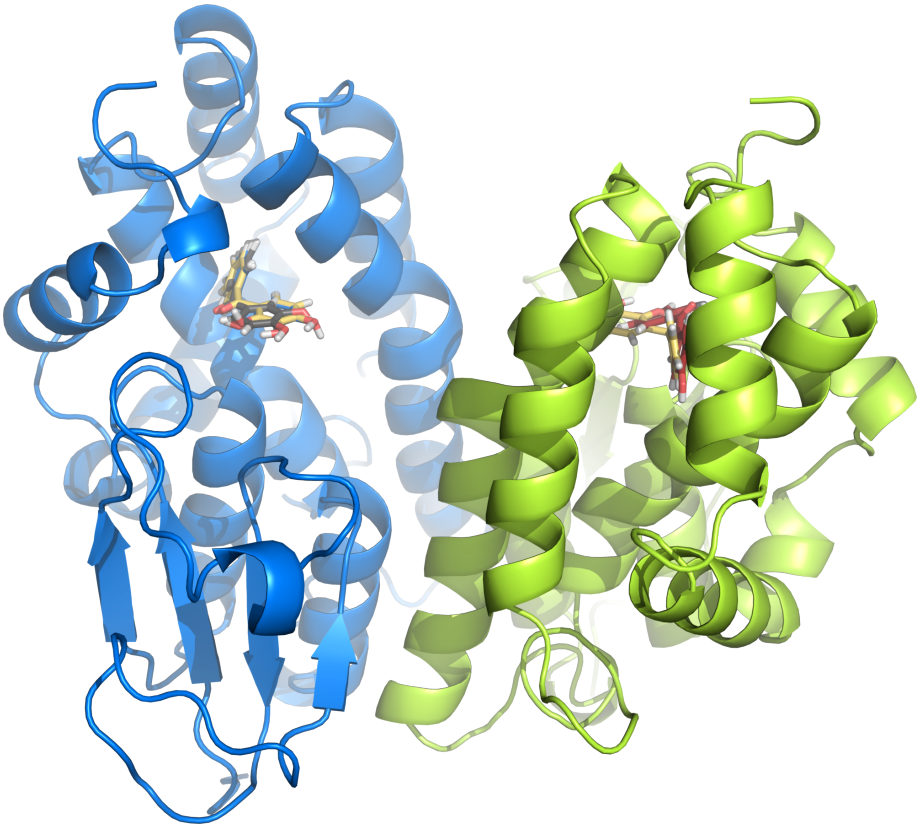


Figure 5.1. The structure of the glutathione transferase (GST) Ω3S enzyme from *Trametes versicolor* shown as a cartoon depicting secondary structures (helices and sheets). The hydroxybenzophenone ligands appear in their binding sites (several ligands in overlay). This GST is a homodimeric enzyme, and the two monomers are shown in different colours (monomer A in blue and monomer B in green).

The particular GST under scrutiny comes from the fungus *Trametes versicolor*.[3] This fungus has the ability to degrade and mineralize all the woods components. Its complex detoxification system copes with secondary metabolites produced by plants. Consequently, it has a considerable arsenal of degradation enzymes, including the Ω class of GSTs (16 Ω-GST genes among 45 GST genes). The Ω3S isoform of glutathione transferase from *Trametes versicolor* (*Tv*GSTO3S, figure 5.1) plays a key role in phase II of enzymic detoxification. It contains a serine as catalytic residue. Recent investigations demonstrate that fungal GSTs, like *Tv*GSTO3S, could be involved in the transport and the sequestration of plant polyphenols.[4]

Schwartz *et al.* determined four structures of TvGSTO3S, each containing a different benzophenone ligand in its active site.[5] Benzophenones are small compounds containing two phenyl rings bound by a carbonyl (C = O) group. Benzophenones mimic the natural ligands of the GST enzyme. The specific benzophenones in question are two di- and two trihydroxybenzophenones (HBPs). Three of them have their hydroxy groups distributed over one phenyl ring (rhe 4'-hydrogen ligands), and the last (2,4,4'-trihydroxybenzophenone, 2,4,4'-HBP) has the hydroxy substituents on both rings (4'-

[3]Linnæan classification: *Fungi*, *Basidiomycota* (higher fungi), *Agaricomycetes*, *Polyporales*, *Polyporaceae*, *Trametes versicolor* (L.) Lloyd (1920).

[4]Schwartz *et al.*, 2018; Perrot, Schwartz, Saiag *et al.*, 2018; Perrot, Schwartz, Deroy *et al.*, 2021.

[5]Schwartz *et al.*, 2018.

hydroxy ligand, table A.3). One can observe a difference in binding conformations of 4'-hydrogen and 4'-hydroxy ligands (figure 5.1).

Some of these structures also contained the glutathione (GSH) ligand in its binding pocket. Glutathione and hydroxybenzophenone substrates predominantly form electrostatic contacts (such as hydrogen bonds), making their complexes with *Tv*GSTO3S textbook examples for a proof-of-concept study that one can reliably apply `Charger` to protein-ligand complexes and obtain good results.

Schwartz *et al.* measured thermodynamic parameters on *Tv*GSTO3S/HBP and *Tv*GSTO3S/GSH systems in solution.[6] A simple parameter is the denaturation temperature T_d, the temperature at which half of the protein in solution unfolds (denaturates from its natural, folded state, figure 5.1). The denaturation temperatures shift in the presence of ligands, usually to higher values,[7] indicating that the ligand presence stabilises the protein-ligand complex. This shift in denaturation temperature is called the thermal shift (ΔT_d). The solution of *Tv*GSTO3S and HBP shows a small, but consistent thermal shift, distinct for various HBP ligands. The magnitude of thermal shifts with GSH ligands is much larger.

One can reliably convert thermal shift values to a dissociation constant K_d that determines the equilibrium state of the chemical reaction *Tv*GSTO$_3$S/HBP $\rightleftharpoons$ *Tv*GSTO$_3$S + HBP. This constant is related to the difference in standard state Gibbs energy (ΔG^o).[8] `Charger` can estimate the electrostatic part of these binding energies: this reveals a path to compare `Charger` calculations directly with experimental results.

One more direction to establish the pertinence of `Charger` calculations is to try and improve protein-ligand binding. The conventional approach seeks to improve strong contacts between them.[9] The bio-macromolecular module connected to `Charger` (`Charger` Protein) can point out residues that improve ligand binding, but also those that hinder ligand binding by poisoning the electrostatic environment of the ligand in the binding site. An *in-silico* inquiry into this topic was pursued from two different directions: `Charger` electrostatic interaction energy calculations and topography of electric field lines. The first approach helped to filter irrelevant residues and focus on the more pertinent ones, while the second clarified what makes those residues pertinent.

5.3 Structure Preparation for Calculations

This section recapitulates the main methodological points used in the paper. More details are given in the methodology section of the article *A rush to explore protein-ligand electrostatic interaction energy with* `Charger` (reproduced in the appendix, under chapter A).

Schwartz *et al.* assessed eight HBPs and obtained crystal structures of four *Tv*GSTO3S/HBP complexes (PDB IDs 6f66, 6f67, 6f68 and 6f69). The PDB files required preparation in `PyMOL` (preliminary

[6] Ibid.
[7] Cimmperman *et al.*, 2008.
[8] Levine, 2009, pp. 161, 162.
[9] Thapa and Raghavachari, 2019.

adjustments), WHAT IF (adding missing hydrogen atoms on amino acid residues and optimising their hydrogen bond networks), MoPro (adding missing ligand hydrogen atoms)[10] as well as manual adjustments of ligand torsion angles to achieve realistic hydrogen bond networks between the ligand and active site residues. Import2MoPro converted the structure into a MoPro-compatible format.[11]

A multipolar database transfer tool exists inside MoProViewer.[12] This tool can transfer parameters automatically onto protein structures, assign charges[13] and transferred experimental multipolar electron density parameters from the ELMAM2 database.[14]

Transferred electron density complemented with theoretical polarisabilities (per atom type, averaged) enables the calculation of polarisation energies. A custom database of theoretical average atomic polarisabilities for ELMAM atoms found in HBPs and proteins was derived. The polarisation methodology followed the method Theo Leduc described in his PhD thesis. The method polarised hydroxybenzophenones and the GST residues around HBP binding site.[15]

*Tv*GSTO3S/benzophenone charge densities in hand permit energy calculations. In all these calculations with Charger, analytical EP was used in the close distance domain up to 5 Å, and pseudo-Buckingham multipolar moments calculation for all remaining atom pairs. The previous chapter (in section 3.5) brought forward some of these calculation results from the perspective of calculation performance, including the use of both cut-off schemes from table A.1. The following section revisits those calculations while keeping an eye for interpretation and the biochemical context. The cited energy values refer to Charger calculations with precise cut-offs.

5.4 Protein-Ligand Interactions

*Tv*GSTO3S is a homodimeric enzyme. GST monomers contain at least two ligand binding sites, the glutathione binding site (G site) and the hydrophobic substrate binding site (H site).[16] Glutathione ligands reside in the G-site in the structures 6f67 and 6f69.[17]

One can find the structures of HBP ligands in the table A.3. They sit in the H-site, and 2,4,4'-HBP (the 4'-hydroxy ligand) has a distinct conformation from that of 2,4-, 2,3,4- and 3,4-HBPs (the 4'-hydrogen ligands) (figures 5.2 and A.4). The conformational difference between the 4'-hydroxy (figure 5.2b) and the 4'-hydrogen (figure 5.2a) ligands lies in the dihedral angles between the car-

[10] Schrödinger, LLC and DeLano, 2020; Hooft, Sander and Vriend, 1996; Jelsch *et al.*, 2005.

[11] Jelsch *et al.*, 2005.

[12] Guillot *et al.*, 2014.

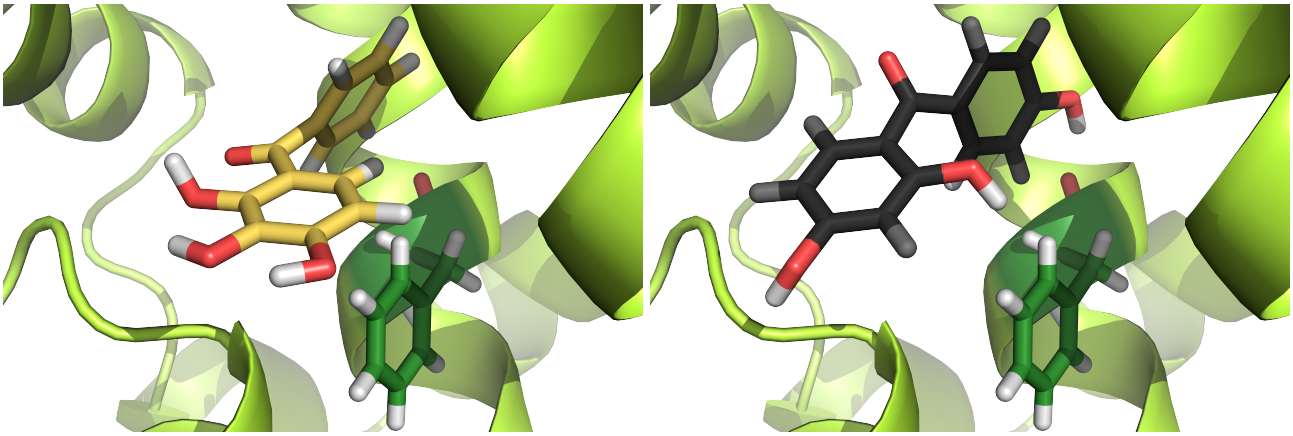
[13] The tool gives the freedom to choose any multiple of elementary charge e to assign to all residues of the same type. For this investigation, the tool distributed $+1\,e$ charge to atoms of arginine and lysine, $-1\,e$ to aspartate and glutamate, $0\,e$ to other residues. The only histidine residue His19 is uncharged.

[14] Domagała *et al.*, 2012.

[15] Leduc *et al.*, 2019; Leduc, 2019.

[16] Schwartz *et al.*, 2018.

[17] The bio-structuralists believe that the glutathiones exist in all four structures, but that the collected data was not sufficiently good to properly resolve them, or that they occupy a small amount of active sites and are thus not sufficiently represented.



(a) A 4'-hydrogen ligand 2,3,4-HBP, from the 6f69 structure.

(b) A 4'-hydroxy ligand 2,4,4'-HBP, from the 6f68 structure.

Figure 5.2. The two different conformations of hydroxybenzophenone (HBP) ligands in the GST active site, related to whether the 4'-substituent is a hydrogen or a hydroxy group. The amino acid residue shown in green is phenylalanine 168, shown to facilitate orientation in the active site.

bonyl plane and the aromatic rings.[18] Thermal shift (ΔT_d) assays show that the 4'-hydrogen HBPs had greater thermal stability when bound to GST compared to 2,4,4'-HBP (the 4'-hydroxy ligand). Schwartz *et al.* did not determine quantitatively how ligand binding correlates with thermal shifts.[19]

`Charger` calculated total electrostatic interaction and polarisation energies between the two subunits and any available ligands. The calculations cover all possible pairs (first monomer – first ligand, first monomer – second ligand *etc.*). Interestingly, `Charger` energies for monomers and their own ligands correlate well with thermal shift ΔT_d (correlation coefficient values generally above 85%, figure A.5, further information in supplementary figure A.S2 and notes under supplementary table A.S6, section A.7).

Furthermore, the order of ΔT_d followed the order of `Charger` energies nearly exactly. The slopes of trend lines ascertained from these four ligands (around −6.5 kcal/mol/K for a monomer and −12 kcal/mol/K for the dimer) indicate that one should predict a thermal shift ΔT_d increase by at least 1 ℃ for the case where the `Charger` energy differs by around −6 kcal/mol/monomer.

These numbers increase slightly when calculating electrostatic and polarisation energy (around −9 kcal/mol/K for a monomer and −17 kcal/mol/K for the dimer). If one produces a small *in silico* modification to the structure (a mutation of a residue, or a slight chemical modification on a ligand, for a couple of examples), and this change results in roughly −6 or −9 kcal/mol energy difference per monomer, one should expect to observe a thermal shift of around 1 ℃. This could serve as a rough

[18] Cox, Kechagias and Kelly, 2008 found both conformations – the 4'-hydroxy and the 4'-hydrogen – for this family of molecules in the solid state. The 4'-hydroxy conformation has the dihedral angle between the aromatic ring at the bottom of the cavity and the carbonyl plane at 30°, while the dihedral angle between the aromatic ring exposed to the solvent and the carbonyl plane is 63°. It is opposite for the 4'-hydrogen conformation, which has the corresponding angles around 55° and 25°.

[19] Schwartz *et al.*, 2018.

guide to select interesting experimental systems based on a ligand design study (*e.g.* with docking) or a mutation study.

5.5 Residue-Ligand Interactions

One often peruses the three-dimensional structure of a macromolecular complex for assumed interactions that favour the adhesion between the macromolecule and the small molecule. This "manual analysis" involves visual inspection and distances measurements. Schwartz *et al.* performed this kind of analysis for the *Tv*GSTO3S/HBP complexes. They found polar (Tyr17 and Arg124), aromatic (Phe123, Trp127, Phe128, Phe168 and Tyr175) and aliphatic groups (the aliphatic part of Arg171) as forming the H-site of *Tv*GSTO3S (figure A.6) and assumed that they accommodate HBPs.

`Charger` calculates the contribution of each individual residue to the total interaction energy between a HBP ligand and *Tv*GSTO3S. These calculations confirm that the monomers are largely similar, a fact also revealed by the superimposition of the subunits (mean RMSD of 0.12 Å). Furthermore, the results for all pairs of monomers hosting 4'-hydrogen ligands (*i.e.* all ligands, except the 2,4,4'-HBP, 6f68) correlate well with each other. The difference between 4'-hydroxy and 4'-hydrogen ligands thus shows up not only as a different conformation and through the lowest thermal shift among the investigated ligands, but also in the energy analysis.

Most residues contribute slightly to the total electrostatic interaction energy (figure 5.3). Few residues contribute significantly, and the most significant interactions surpass ±3 kcal/mol (figure A.6). Overall, they fall into four categories:

- residues suspected by visual inspection and found by `Charger` calculations, which favour HBP binding (Tyr17, Arg124, Trp127, Phe128, Phe168 and Tyr175),
- residues found from `Charger` calculations, but not through a visual inspection, which favour HBP binding (Pro16 and Lys55),
- residues assumed to be important through visual inspection, but not found using the `Charger` calculation (Phe123),
- residues assumed through a visual inspection, and found using `Charger` calculations, which favour binding of some, and disfavour binding of other HBPs (Arg20, Asn120, Arg171).

The article (found in appendix A) examines two residues from the first group (Tyr17 and Phe168, showing typical hydrogen bonding and C-H···π interactions patterns detected by `Charger`) and one from the second group (Pro16, presumably due to close contacts). This displays `Charger`'s sensitivity towards typical electrostatic interactions in biological systems. The residue Phe168 stabilises the particular conformation of 2,4,4'-HBP because of an additional hydrogen bond with the 4'-hydroxy group, unrealisable with other ligands.

The last category of residues opens a distinct opportunity. In particular, one would assume that Arg171 enjoys weak but favourable C-H···π and van der Waals interactions with HBP ligands.[20] A

[20] Schwartz *et al.*, 2018.

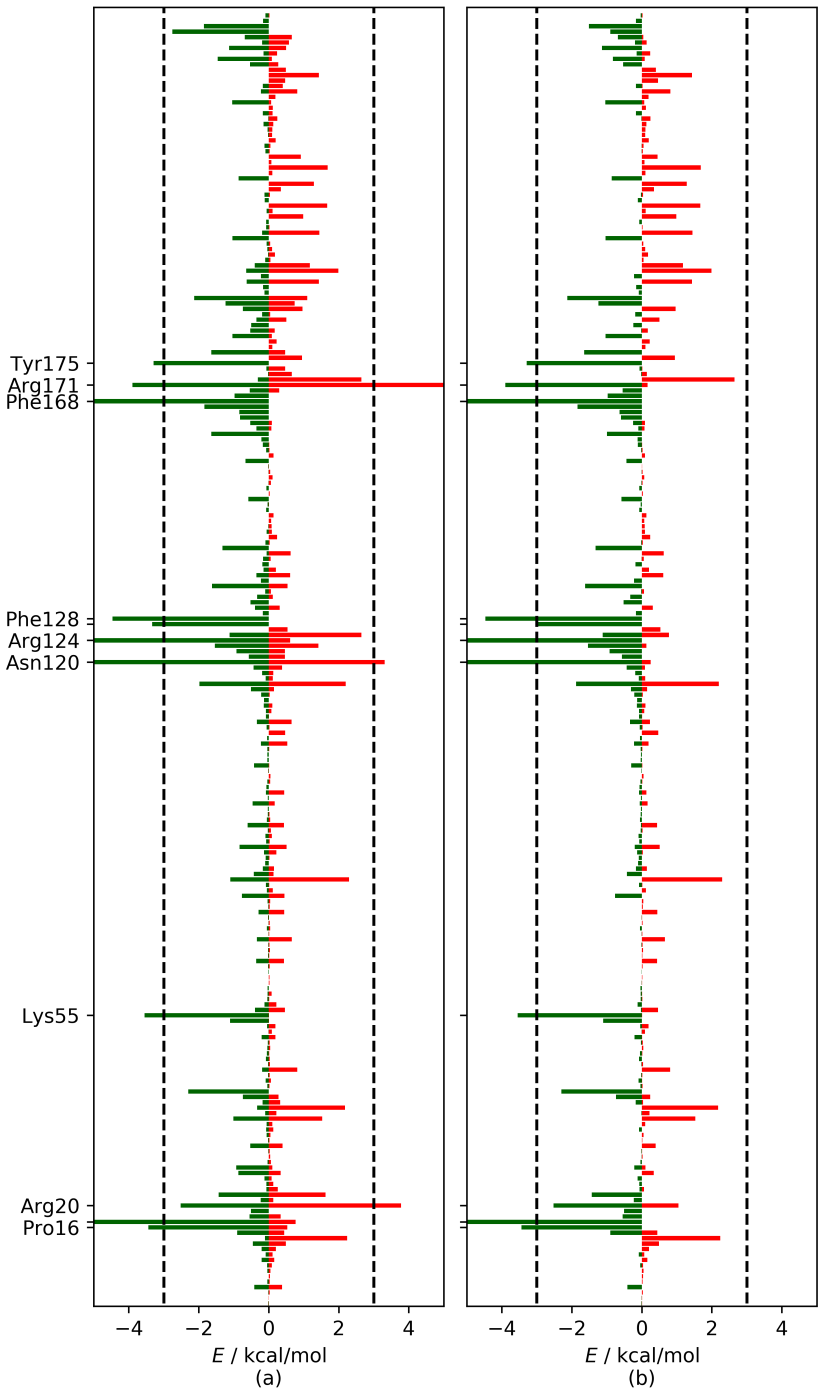


Figure 5.3. Largest favourable (green) and unfavourable (red) electrostatic interaction energy contributions for individual GST residues and the benzophenone ligands. For each residue, the largest favourable energy with all four ligands, as well as the largest unfavourable energy for all four ligands is retained. Energies have been capped at ±5 kcal/mol. Ticks on the y-axis represent the residues with at least one interaction energy contribution beyond ±3 kcal/mol (dashed black lines). The omitted labels are Tyr17 and Trp127. (a) All four benzophenone ligands are considered. (b) Only the three benzophenone ligands with a similar binding mode are considered. Taken together, these graphs show that many amino acids contribute a small amount to binding from the electrostatic point of view, but their contributions roughly balance out. The energy contributions of residues outside the ±3 kcal/mol are highlighted on figure A.6. This figure also exists in the supplementary material of the paper *A rush to explore protein-ligand electrostatic interaction energy with* `Charger`.

different interpretation appears under "`Charger` 's microscope," though one should keep in mind that `Charger` can not estimate van der Waals interactions. As expected, it predicts a strong favourable interaction energy between Arg171 and 4-hydrogen HBPs (-2.6 ± 1.1 kcal/mol, mean ± s.d.). Surprisingly, the interaction energy between Arg171 and 2,4,4'-HBP is unfavourable by about +4.5 kcal/mol (figure A.6).

The estimated electrostatic and polarisation energy also supports the general conclusion. The polarisation largely compensates for the negative effect on 2,4,4'-HBP binding (electrostatic interaction energy reduces to +0.8 kcal/mol), and increases the binding propensity of other ligands (interaction energy of -7.9 ± 1.6 kcal/mol, mean ± s.d.). Arg171 does not hinder 2,4,4'-HBP binding so severely as one could conclude without polarisation density.

Some explanations on these strange binding effects by Arg171 come from molecular dipole moment analysis. `VMoPro` calculates them, and `MoProViewer` displays the results. The dipole moments of 2,4,4'-HBP and Arg171 are at angles of roughly 60° to 90° in the two monomers (figure A.7.a), in such a way that the positive charge distributions of Arg171 and 2,4,4'-HBP seem to interfere. On the other hand, the dipole moment of 4-hydrogen HBPs and that of Arg171 are anti-parallel (angles of around 150 to 170°, figure A.7.b). This could signal a favourable Keesom (dipole-dipole) interaction.

5.6 The Arg171Gln Mutant

The section *Protein-Ligand Interactions* concluded by postulating that a small difference of electrostatic interaction energy, roughly −6 or −9 kcal/mol/monomer (non-polarised / polarised respectively), in an *in silico* mutant should produce an observable thermal shift ΔT_d difference of about 1 ℃. The section *Residue-Ligand Interactions* discussed how arginine 171 has a negative impact on 2,4,4'-HBP binding because of their charge distributions. This section puts these two pieces together.

The arginine charge appears as the biggest roadblock for a favourable interaction. An attenuation of the arginine charge could determine whether charge and polarisation are responsible for the observed (weakly unfavourable) electrostatic binding energy. Glutamine seemed suitable mutation candidate, with its fairly electronegative oxygen atom, the lack of formal charge, and being only slightly less bulky than an arginine.

The *in silico* Arg171Gln mutation took place on the 6f68 structure in PyMOL.[21] The multipolar parameter transfer happened in the same way as for non-mutated structures.

The analysis revealed that the Gln171 dipole moment is indeed much smaller than that of Arg171. The dipole orientation relative to 2,4,4'-HBP persisted (figure A.7.c,d). However, the negative effect disappears when passing from the wild-type Arg171 (+4.5 kcal/mol) to the mutant Gln171 (−0.2 kcal/mol).

Moreover, polarisation density revealed that the Arg171Gln mutation affects the entire binding pocket. The polarised Arg171 wild-type is still unfavourable to binding (+0.8 kcal/mol), but the Gln171

[21] Schrödinger, LLC and DeLano, 2020.

mutant becomes clearly favourable after polarisation (−2.2 kcal/mol). Glutamate 171 is not charged, so its diminished dipole moment would contribute to this favourable interaction.

The developments on the single amino acid energy appear promising, but the real question for an experimentalist lies on a larger scale. The electrostatic and polarisation interaction energies with 2,4,4'-HBP do indeed improve when moving from the wild-type Arg171 (−41.4 and −84.0 kcal/mol, monomer and dimer) to the Gln171 mutant (−50.3 and −101.6 kcal/mol, monomer and dimer). This corresponds to approximately 1 ℃ difference in thermal shift ΔT_d, making Arg171Gln an attractive mutation candidate. One could not reach this conclusion so easily from non-polarised energy calculations: this fact emphasises the importance of polarisation corrections.

5.7 Electric Field Lines Topography

Many thanks to T. Leduc for his help and discussions about this section.

The previous section provided a rationale for embarking on experimental evaluation based on electrostatic interaction energy calculations. This section turns towards electric field lines in order to rationalise the improvement in the microcosm of the 2,4,4'-trihydroxybenzophenone binding site.

Since the 2,4,4'-HBP ligand is neutral, the effect that the loss of the arginine positive charge has on binding should not be large. On top of that, dipole moments of the ligand and of the arginine 171 are nearly perpendicular, but permanent electrostatic interaction between 2,4,4'-HBP and Arg171 is not favourable (+4.5 kcal/mol). The polarisation energy compensates that handicap (electrostatic and polarisation energy of +0.8 kcal/mol), though molecular dipole moments do not change drastically. The electrostatic interaction energy in the Arg171Gln mutant between Gln171 and 2,4,4'-HBP favours the binding (+0.8 kcal/mol), and adding polarisation terms (−2.1 kcal/mol) improve this interaction even more.

`MoProViewer` has a facility to plot electric field lines in the protein active site.[22] This helped to pinpoint the main sources of electrostatic influence in the 6f66 and 6f68 structures and in the Arg171Gln mutant derived from 6f68. Ligand polarisation comes from electrostatic influence of active site residues, so this procedure brings forward those residues that mainly influence the bound ligand.

The electrostatic influence zone linked to the residue 171 remains fairly conserved in the chains A and B. One would expect this from similarities between the chains discussed in section *Residue-Ligand Interactions*. Similar zones persist in both the 6f66 (with 4'-hydrogen HBP) and 6f68 (with 4'-hydroxy HBP) complexes, despite the different ligand binding modes.

In the wild type 6f68 (with 4'-hydroxy HBP, figure 5.4.A), the dark red bundle starts from the C_δ atom of Arg171 and converges to the negatively charged backbone oxygen of Phe168. This bundle

[22] Guillot *et al.*, 2014.

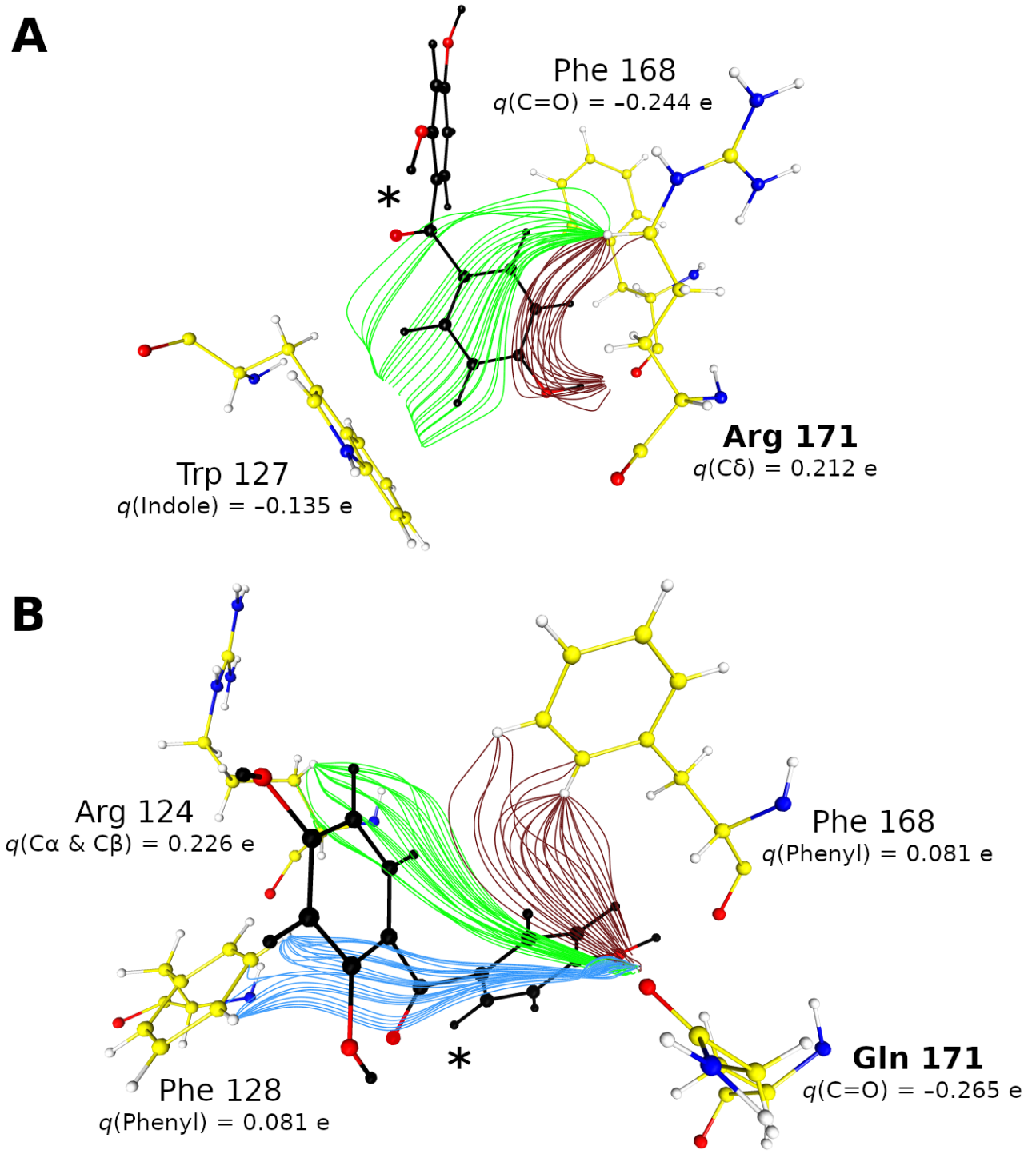


Figure 5.4. **(A)** Electric field lines converging to the main chain of Arg171 in the active site of the wild-type GST (chain B, PDB code: 6f68) bound to 2,4,4'-trihydroxybenzophenone ligand (marked by an asterisk). The dark red bundle originates from the δ-methylene of Arg171 and terminates on the lone electron pairs of the backbone oxygen of Phe168. The green bundle starts from the δ-methylene of Arg171 and ends on the π-electrons of Trp127. Charges are displayed for involved functional groups to give a sense of how steep the electric field is in a given bundle. **(B)** Electric field lines converging to the lone pairs of the Gln171 δ-oxygen, in the active site of the Arg171Gln GST variant (chain B), bound to the 2,4,4'-trihydroxybenzophenone ligand (marked by an asterisk). The dark red bundle starts on the phenyl hydrogen atoms of Phe168. The green bundle starts on the alpha and beta positions of Arg124, covering atoms from both benzophenone cycles. The dark blue bundle starts on phenyl hydrogen atoms from Phe128.

covers two carbon atoms from the 2,4,4'-HBP aromatic ring C1'–C6' and the 4'-hydroxy group. The latter is specifically involved in the strong hydrogen bond with the Phe168 backbone oxygen atom.

The green bundle in the wild type converges to the π-electron cloud of Trp127 (figure 5.4.A). It covers the remainder of the ligand aromatic ring C1'–C6' and the keto group that bridges the phenol rings. Arg171 does not seem to have a major electrostatic influence beyond the keto group in the wild type. The green field lines explain the direction of induced dipole moments in a large portion of the ligand. They align reasonably with the ring plane, and thus with the polarisabilities of the aromatic carbon atoms (figure 5.4.A). This could explain the reduction of the unfavourable electrostatic interaction energy between Arg171 and the ligand.

Going from the wild type (figure 5.4.A) to the Arg171Gln mutant (figure 5.4.B), one should keep in mind that the field lines run in opposite directions. The residues Phe14, Pro16 and Tyr175 are in close proximity to the residue 171. Bundles of field lines join these residues to the residue 171 in the wild type and in the mutated protein, but the considered bundles do not cover ligand atomic positions (and are omitted from figure 5.4).

In the mutant, three bundles cover a large part of the 4'-hydroxy ligand and end at the Gln171 side chain. They originate from three different residues: Arg124, Phe128 and Phe168. The electrostatic influence of Gln171 seems thus more important than that of the Arg171 in the wild type. This is in agreement to electrostatic energy results from `Charger`, indicating a mild stabilising influence of Gln171 in the active site. These findings are consistent with interpreting the effect of the Arg171Gln mutation as an electrostatic obstacle removal.

6 LiDHQ

6.1 Résumé

Ce chapitre donne une perspective sur le 2,5-dihydroxyquinonate (DHQ^{2-}) de lithium, co-cristallisé avec deux molécules de l'eau et deux molécules de dimethylsulfoxyde (DMSO). Le cristal a été découvert fortuitement : les cristaux du dihydrate de 2,5-dihydroxyquinonate du lithium étaient trop petits pour une mesure par diffraction de rayons X, et ils ont été recristallisés avec du DMSO qui s'est intercallé dans le cristal.

Un contact proche entre la paire libre du soufre du DMSO et les atomes de carbone du DHQ^{2-} apparaît dans la structure. C'est le premier cas de ce type de contact impliquant un dianion, et il a été soigneusement étudié pour bien établir la nature de ce contact. Une combinaison de la densité de charge expérimentale et de la modélisation théorique provenant de la chimie quantique a révélé la nature du contact.

La topologie de la densité électronique suggère un point critique entre l'atome S sur le DMSO et un atome C carbonyl sur le quinoïde, mais l'analyse du potentiel électrostatique indique que les atomes se repoussent. L'énergie d'interaction globale entre DMSO et DHQ^{2-} est attractive (–19.1 kJ/mol), ainsi que l'énergie totale d'interaction calculée *in vacuo* (–42.8 kJ/mol). De plus, les calculs DFT *in vacuo* dévoilent que cette interaction peut alternativement être considérée comme une liaison chalcogène.

6.2 Introduction

Intermolecular interactions involving electron density depletions of a π-system recently came into research focus. They are generally named π-hole interactions.[1] They typically involve electrostatic attraction between an electron-rich group and an electron-depleted region of a π-electron system.[2] Some explanations are known for such interactions: donation of a lone electron pair (lp) into an empty π* orbital (*i.e.* lone pair···π or n⟶π* interactions),[3] an anion's negative charge forming anion···π interactions.[4]

These interactions form in the presence of a positive electrostatic potential (ESP) zone, or an electron depletion, in a conjugated system (i.e. a π-hole) on the acceptor. One typically observes them in aromatic rings with multiple electron-withdrawing substituents, with an alternative in non-aromatic

[1] Quiñonero *et al.*, 2002; Frontera *et al.*, 2011; Kozuch, 2016; Angarov and Kozuch, 2018.

[2] Kozuch, 2016; Angarov and Kozuch, 2018; Grounds, Zeller and Rosokha, 2018; Jia, Miao and Hay, 2019.

[3] Angarov and Kozuch, 2018; Mooibroek, Gamez and Reedijk, 2008; Singh and Das, 2015; Newberry and Raines, 2017.

[4] Frontera *et al.*, 2011; Schottel, Chifotides and Dunbar, 2008; Gamez *et al.*, 2007; D.-X. Wang and M.-X. Wang, 2013; Lucas *et al.*, 2016; Savastano *et al.*, 2017.

perhalogenoquinones.[5] Generally, the strongly electron-withdrawing carbonyl groups make quinoid rings inherently electron-poor. The presence of π electrons that are much more localised than in the aromatic groups exacerbates the strength of carbonyl groups (*i.e.* single and double bonds in quinones can be easily distinguished).[6] Electronegative substituents, such as halogens, further deplete the electron density in quinoid rings, forming very pronounced π-holes at carbonyl C atoms.[7] Several groups observed intermolecular contacts shorter than the sum of van der Waals radii between perhalogenquinones and halide anions.[8]

Only severely electron-depleted π systems were scrutinised as possible acceptors of π-hole interactions. The paper *An unusual interaction of a lone pair and an π-electrons in a quinoid dianion* (included as appendix B) presents a detailed study of 2,5-dihydroxyquinonate (DHQ^{2-}), an electron-rich dianionic quionid ring, in a serendipitously obtained compound featuring a lone pair···π-hole interaction. This is the first instance of an anionic quinoid ring participating in such an interaction.

The study relied on a combination of experimental X-ray charge density and quantum chemical modelling. A brief methodology recapitulation follows. Then the focus falls squarely on this contact resembling a lone pair···π-hole interaction, and the manner in which `Charger` contributed to its characterisation.

6.3 Experimental

A recrystallisation of lithium 2,5-dihydroxyquinonate crystals from DMSO yielded diaqua*bis*(dimethylsulfoxyde)-μ-(2,5-dihydroxyquinonato)dilithium ($[Li_2\mu\text{-}(C_6H_2O_4)(H_2O)_2((CH_3)_2SO)_2]$, shortened to **LiDHQ**, scheme B.1) crystals serendipitously.[9] Diffraction data were collected on a Bruker D8 Venture diffractometer[10] for a single crystal of **LiDHQ** using MoKα radiation (λ = 0.71073 Å). Data processing was performed with the `Bruker SAINT` software.[11] The resolution of the collected dataset goes to 1.0 $Å^{-1}$, with 99.7% completeness and R_{int} = 7.06%. One can find more detailed statistics in table B.3.[12]

[5] Molčanov, Mali *et al.*, 2018; Kepler, Zeller and Rosokha, 2019; Wilson *et al.*, 2020.

[6] Vuković, Molčanov *et al.*, 2019.

[7] Molčanov, Mali *et al.*, 2018; Molčanov, Jelsch *et al.*, 2018; Molčanov, Milašinović and Kojić-Prodić, 2019; Molčanov and Kojić-Prodić, 2019.

[8] Molčanov, Mali *et al.*, 2018; Kepler, Zeller and Rosokha, 2019; Wilson *et al.*, 2020.

[9] The synthesis of lithium 2,5-dihydroxyquinonate followed the protocols established in the papers by **Xiang2008** and by Molčanov, Kojić-Prodić *et al.*, 2013.

[10] Equipment: a molybdenum micro-focus X-ray tube with a mirror monochromator, and a PHOTON 100 CMOS detector.

[11] Bruker, 2012.

[12] R is the agreement factor, a common measure in crystallography of how much a dataset deviates from theoretical values. The subscript *int* is short for "internal", a measure of internal consistency of the observed reflections.

The Olex2 graphical user interface served to solve and refine the structure of **LiDHQ**.[13] It used external programs SHELXS97 and SHELXT-2017.[14] SORTAV then removed symmetry-forbidden reflections and averaged the remaining ones.[15]

The program MoPro refined the crystal structure using the Hansen-Coppens multipolar model.[16] A gradual refinement of parameters was executed. The following multipolar coefficients were used: none for lithium atoms, up to dipolar level for hydrogen atoms, octupolar level for carbon and oxygen atoms, hexadecapolar level for the sulphur atom.

A calculation on the SHADE3 server determined anisotropic displacement parameters U_{ani}(H) of hydrogen atoms from 2,5-dihydroxyquinonate dianion and dimethylsulfoxyde.[17] Calculated values of U_{ani}(H) were constrained to those from SHADE3 during the refinement. Furthermore, dipolar, quadrupolar and expansion/contraction parameters for the water molecule were constrained to those from the ELMAM2 database.[18] Several *R*-free refinements with 20 refinement cycles determined optimal restraint values for multipolar local symmetry and valence/multipolar population similarities.[19]

The model quality was evaluated through three graphical techniques: the DR plot,[20] the resolution plot,[21] and the fractal dimension plot.[22] XDRKplots made the first two plots, while jnk2RDA created the third.[23] They are all available in the *Supporting Information* (section B.6). The grounds to decide in which direction to proceed with the refinement were a compromise between improving these plots, residual density global minimum/maximum values and deformation density contour plots.

VMoPro calculated total charge density (3D grids, in Gaussian cube format) for the entire complex (two asymmetric units related by inversion in the centre of the quinone ring). VMoPro further calculated ESPs, bond critical points and their parameters (density, ellipticity, density Laplacian etc.).[24]

Atomic charges come from atomic basin integration using the Critic2 program, which uses three-dimensional density grids as input files.[25] Topological bond orders were determined using an empiri-

[13]Dolomanov *et al.*, 2009.

[14]Sheldrick, 2007; Bourhis *et al.*, 2015b.

[15]Blessing, 1995.

[16]Jelsch *et al.*, 2005.

[17]Madsen, 2006; Madsen and Hoser, 2015.

[18]Domagała *et al.*, 2012.

[19]An *R*-free refinement refines the model against 95% of the dataset and then uses the remaining 5% to calculate the agreement factor *R* and decide whether the model fits the data well. This procedure helps to avoid overfitting.

[20]Abraham and Keve, 1971. The DR plot contains $(I_{\text{obs}} - I_{\text{calc}})/\sigma$ versus theoretical quantiles from a standard normal distribution. The symbol *I* denotes the intensity of a crystallographic reflection, which can be *ob*served (measured) or *calc*ulated from the content and dimensions of the unit cell. The value σ is the standard deviation of the observed reflection.

[21]Zhurov, Zhurova and Pinkerton, 2008. The plot of $I_{\text{obs}}/I_{\text{calc}}$ versus $\sin(\theta)/\lambda$, the reciprocal resolution of a particular reflection. I_{obs} and I_{calc} are explained in the previous footnote.

[22]A residual density isosurface has a fractal dimension associated to it. The fractal dimension plot shows the fractal dimension as a function of the isosurface value. If the residual density after a crystallographic refinement is made up of purely Gaussian noise, the plot will look like a parabola with a maximum at $(0, 3)$. In regular datasets the y value of the maximum is usually between 2.5 and 3. The paraboloid shape is more relevant when evaluating data quality.

[23]Meindl and Henn, 2008; Meindl, 2008.

[24]Jelsch *et al.*, 2005.

[25]Tang, Sanville and Henkelman, 2009.

cal formula.[26] `MoProViewer` created three-dimensional visual representations of deformation density and ESP surfaces.[27] `VMoPro` generated two-dimensional property plots.[28]

Two types of quantum chemistry computations were performed: periodic DFT using `Crystal14`[29] and *in vacuo* DFT using `Gaussian 09`.[30] Periodic DFT calculations started from atomic coordinates after multipolar refinement. These calculations resided on the B3LYP/6-31G(d,p) level of theory adapted for periodic calculations in crystals, and other recommendations from literature were observed.[31]

The program `properties` from the `Crystal14` package derived structural factors up to 1.28 $Å^{-1}$ resolution from the calculated periodic wave-function.[32] Calculated structure factors were used to refine a Hansen-Coppens charge density using `MoPro`. Only charge density-related parameters were refined. The obtained parameters served to restrain physically non-justifiable κ and κ' values in experimentally determined charge densities. These *in crystallo* calculations were only used to compare to and improve experimental charge density models.

The *in vacuo* calculations were used to optimise geometries, determine electronic (not electrostatic) interaction energies on cystal geometries and calculate electron density difference (EDD) maps. The optimisation and energy calculation was performed on the B3LYP/jun-cc-pVTZ level of theory with Grimme's D3 dispersion correction.[33] The EDD calculation operated at the CAM-B3LYP/jun-cc-pVTZ level.[34]

6.4 Interaction between the lone pair of DMSO and the π-system of the DHQ dianion

The centrosymmetric moiety $Li_2DHQ(H_2O)_2(DMSO)_2$ is the elementary building block of the LiDHQ crystal structure, with the DHQ^{2-} ring centroid located at the crystallographic inversion centre (scheme B.1, figure 6.1). The lithium cation lies in a centre of a tetrahedron containing four oxygen atoms at its edges: O1 and O2 atoms from DHQ, DMSO oxygen O3, and water molecule oxygen O4.

Two symmetry-independent O-H···O hydrogen bonds and seven weaker C-H···O bonds exist. They form layers parallel to the $(\bar{1}01)$ crystallographic direction (fig. B.2). There is no cycle stacking of **DHQ** rings in the crystal structure.

A close contact between the DMSO sulphur atom and the DHQ^{2-} ring stands apart as the most intriguing facet of the crystal structure (figure 6.1). The closest separation is between S1···C2 of 3.2481(7) Å, followed by S1···C1 of 3.4046(7) Å; the distance between S1 and the DHQ^{2-} ring cen-

[26] Howard and Lamarche, 2003; Tsirelson *et al.*, 2007; Bartashevich *et al.*, 2011.
[27] Guillot *et al.*, 2014.
[28] Jelsch *et al.*, 2005.
[29] Dovesi, Orlando *et al.*, 2014; Dovesi, Saunders *et al.*, 2014.
[30] Frisch *et al.*, 2009.
[31] Gatti, Saunders and Roetti, 1994; Heyd *et al.*, 2005; Clark *et al.*, 1983.
[32] Dovesi, Orlando *et al.*, 2014; Dovesi, Saunders *et al.*, 2014.
[33] Becke, 1993 gives the functional, Papajak *et al.*, 2011 the basis set, and Grimme *et al.*, 2010 the dispersion correction.
[34] Yanai, Tew and Handy, 2004 describe the functional, and Papajak *et al.*, 2011 the basis set.

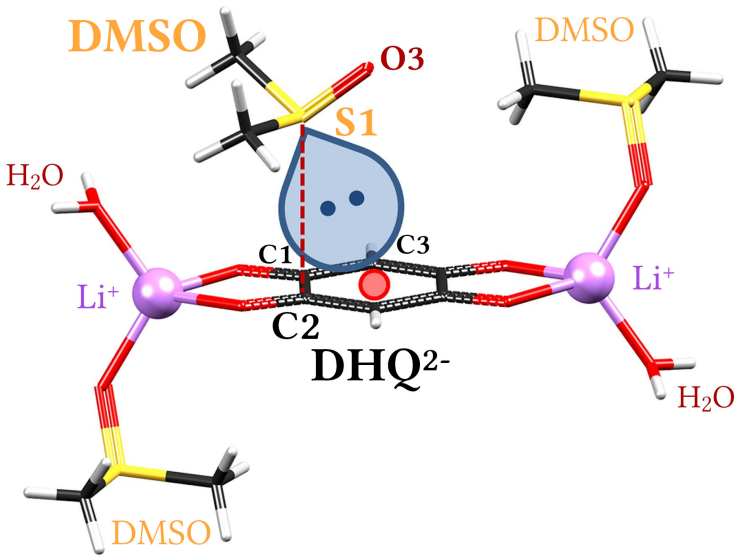


Figure 6.1. The close contact in the serendipitous structure of LiDHQ.

tre is 3.461 Å. The lone electron pair of S1 points towards the π-system of the quinoid ring (figure B.3). One would not expect such an interaction to be electrostatically attractive.

In fact, the DMSO generates an ESP of −0.05 V at the sulphur lone pair location, while DHQ^{2-} generates an ESP of −0.55 V at the π-hole location generated by the atom C2 on the Hirshfeld surface. The negative ESP surrounding the electron-depleted C2 atom results from the global negative charge of the DHQ^{2-} dianion. Moreover, the q_{model} atomic charge derived from P_{val} is $+0.41e$ for S1, while on the DHQ^{2-} side it is $0.00e$ for C1 and $+0.01e$ for C2 – the π-hole atoms appear richer in electrons than the S1 sulphur (figures B.1.a and B.5). This interaction appears opposite of typical π-hole interaction. From a purely electrostatic point of view, it seems to be some kind of an inverse π-hole interaction.

The new Charger electrostatic interaction energy tool broadens this simple charge-based analysis.[35] It considers all charge density descriptors around chosen pseudo-atoms. Both S1···C1 and S1···C2 electrostatic interaction energies are slightly positive (+0.9 and +2.5 kJ/mol, respectively): the pseudo-atoms involved in direct contact push each other away. However, considering the S1=O3 group in the place of the S1 atom makes the interaction attractive: electrostatic contact energies (O3=S1)···C1 and (O3=S1)···C2 become −0.5 and −1.2 kJ/mol, respectively. The electronegative oxygen O3 presumably compensates the observed repulsion between sulphur and carbon atoms. This shows up often when working with a multipolar model, since it is the ensemble of multipolar parameters on two pseudo-atoms that form a covalent bond that describes its density.

However, the interaction between S1 atom and the two C3 atoms in the quinoid ring is strongly attractive (electrostatic interaction energy of −36.2 and −32.8 kJ/mol), in spite of large contact distances (3.6386(7) and 3.8371(7) Å). The largest contribution to this attractive energy comes from the $N_{\text{val}} - P_{\text{val}}$ charges on the sulphur S1 ($+0.41e$) and carbon C3 ($-0.24e$) bound to a hydrogen atom.

[35] Vuković, Leduc *et al.*, 2021.

Furthermore, the overall DMSO···DHQ^{2-} interaction energy (−19.1 kJ/mol) indicates a strong global electrostatic attraction between them.

The topology of electron density has many (3, −1) ("bonding") critical points (CP, table B.2, figure B.6). The CPs with largest electron density correspond to O-H···O hydrogen bonds, followed by the S1···C1 contact ($\rho_{CP} = 0.041e\,Å^{-3}$), and C-H···O hydrogen bonds are considerably weaker. Contrary to their name, the presence of (3, −1) critical points does not always imply the existence of attractive interactions (bonds),[36] especially when dealing with weak intermolecular interactions.[37] Charges and `Charger` analyses indicate that the S1···C1 atom contact is not attractive from a pure electrostatic point of view. Instead, it results from the global attraction between the DHQ^{2-} and DMSO entities.

One must also consider quantum-chemical (DFT) *in vacuo* computations, which tell a different tale. These calculations were separate from the *in crystallo* calculations and were not used for parameter refinement. A geometry optimization of a DMSO···DHQ^{2-} complex *in vacuo*, starting from the geometry found in the crystal structure, did not result in a complex with either an S···C or an S···O interaction. Instead, it yielded a C-H···O hydrogen-bonded complex with a DMSO methyl group as a donor.

The total (electronic, not electrostatic) energy of interaction between DHQ^{2-} and DMSO at the crystal-structure geometry of DMSO···DHQ^{2-} is indeed strongly attractive (−42.8 kJ/mol). The electron density difference (EDD) upon formation of the interaction (figure B.7) shows that the interaction moves electron density from the lone-pair region of S1 closer to its nucleus and to the atom O1 of DHQ^{2-}.

6.5 Conclusions

The crystal structure of **LiDHQ** contains an unusual intermolecular contact resembling a π-hole interaction involving a dianion as an acceptor. The interaction itself is rather confusing, so it was approached from various theoretical and experimental perspectives.

The electron density topology analysis (the Bader QTAIM approach) indicated the presence of a critical point between S1 and C2 pseudo-atoms. The ESP showed that the interaction cannot be localised, and that the electrostatic interaction between pseudo-atoms S1 and C2 directly involved in this interaction must be in fact slightly repulsive. On the contrary, the entire DMSO molecule and the DHQ^{2-} dianion attract each other with an electrostatic energy reaching −19.1 kJ/mol.

Analysis of ESP and `Charger` electrostatic energy calculations showed a strong attraction between $S1^{\delta+}$ and $C3^{\delta-}$ atoms. Therefore, no local attraction between S1 and C2 atoms exist, despite the presence of a bonding CP, but that the overall interaction between DHQ^{2-} and DMSO is attractive. It contains an electrostatic and a dispersive component. The presence of the CP thus results from a global attractive interaction between DHQ^{2-} and DMSO, as well as a combination of other interactions present in the crystal field.

[36] Shahbazian, 2018; Bader, 2009.
[37] Wick and Clark, 2018.

The use of `Charger` improved the understanding by providing further proof that the S1 and C2 pseudo-atoms repel each other, and pointed to an important role of the C3 atom. The total electronic energy estimated from an *in vacuo* DFT calculation (−42.8 kJ/mol) is double that of `Charger` electrostatic interaction energy (−19.1 kJ/mol). This should caution the users against relying on `Charger` alone to predict total interaction energy. The DMSO···DHQ^{2-} interaction is thus both electrostatic and dispersive in nature. This gives an incentive for the research into a bridge from electrostatic energy, over the van der Waals chasm, to the total interaction energy of an investigated system described with a multipolar electron density from a Hansen-Coppens model.

7 Cyclopentadienyl‴NiP_3 and Crystal Electrostatic Energy

Many thanks to F. Meurer for his help and discussions about this chapter, in particular the parts about Cp‴NiP_3.

7.1 Résumé

Ce chapitre lie deux études. La première est sur un complexe organométallique de carbone, hydrogène, nickel et phosphore : Cp‴Ni(μ_3-*cyclo*-P_3), Cp‴ étant l'anion 1,2,4-tris(*tert*-butyl)-cyclopentadienyl). Il imite le phosphore blanc (P_4) et peut substituer les précurseurs toxiques utilisés dans la synthèse industrielle.

Comme dans le chapitre précédent, un modèle multipolaire a été soigneusement construit à partir des données de diffraction des rayons X collectées dans le cadre d'une collaboration Franco-Allemande. Contrairement au cristal du chapitre précédent, il n'y avait pas beaucoup de contacts intermoléculaires forts. La chimie du complexe est fascinante : les liaisons tordues entre les atomes du phosphore (figure 7.1b), les charges des diverses parties du complexe (table 7.2) et l'analyse des orbitales *d* sur le nickel (table 7.2). `Charger` a révélé les fortes interactions électrostatiques entre les trois parties du complexe.

La deuxième étude est déjà largement commentée dans un précédent article.[1] Il s'agit du premier composé que l'auteur a cristallisé et modélisé avec le modèle multipolaire : l'acide chloranilique (schéma 7.1). Ce chapitre n'aborde que quelques détails de la grande et riche entreprise d'investigation de cette molécule singulière.

Le pont entre les deux études provient d'une nouvelle application de `Charger` : le calcul de l'énergie électrostatique de la maille cristalline. `Charger`, étant limité au calcul sur des ensembles finis, ne peut pas répondre à ce défi directement. Il approche quand même un résultat fiable dans la mesure du possible (figure 7.5).

[1] Vuković, Molčanov *et al.*, 2019.

7.2 Cyclopentadienyl‴NiP_3

In the industrial world, the standard synthesis of organo-phosphorous compounds requires chlorination of white phosphorous (P_4) to phosphorous trichloride (PCl_3). One functionalises PCl_3 through the use of a reductive agent (alkali metals) and either Grignard/organolithium compounds or organochlorides. The major drawbacks of this synthesis pathway come from the toxic and environmentally hazardous compounds used (notably the chlorine gas (Cl_2) and phosphorous trichloride) as well as ionic salts formed in the neutralisation.

One attractive alternative is the activation of P_4 mediated by a transition metal. The reactivity of phosphorous enhances through the introduction of an organometallic fragment, such as CpM (Cp = Cyclopentadienyl or a derivative, M = transition metal).

One exemplary synthesis precursor is the titular compound abbreviated to Cp‴NiP_3 (more specifically, Cp‴Ni(μ_3-*cyclo*-P_3), Cp‴ = 1,2,4-tris(*tert*-butyl)-cyclopentadienyl anion). The isolobality concept[2] permits to view the nickel fragment (Cp‴Ni) as isolobal to one phosphorous atom, yielding a compound isolobal to white phosphorous (P_4).[3]

This view opens up interesting opportunities for functionalization: for one example, the *cyclo*-P_3 fragment can undergo P-P bond insertion by *in situ*-generated phosphenium ions $[RR'P]^+$. The cationic *cyclo*-P_4 fragment generated in this insertion of electrophilic moieties remains coordinated to the Ni metal fragment.[4] The organometallic chemistry of Cp‴NiP_3 is also rich and worth exploring.[5]

Another benefit of using Cp‴NiP_3 is its stability in air. Despite this, it shows a rich redox chemistry and enhances the possibilities of organo-phosphorous synthesis. Open questions linger regarding its electronic structure, especially at the metal centre, in spite of its many applications. Nobody has thus far devoted time to realise state-of-the-art calculations on Cp‴NiP_3 to the best of author's knowledge. These calculations, coupled with experimental charge density data, promise to explain observed reactivity of the compound. The focus here will rest on `Charger` calculations, but the time approaches when a large endeavour to survey this marvellous compound shall culminate.

[2]Isolobality, defined in the paper Elian *et al.*, 1976, looks at the shape and energy of frontier orbitals and to similar electron counts in order to draw analogies between well-studied chemical systems and those on which less information is available. A Nobel prize was awarded for this concept in 1982, and the award speech that further clarifies this contact can be found in the paper by Hoffmann, 1982.

[3]Scheer, Balázs and Seitz, 2010; Cossairt, Piro and Cummins, 2010; Hoidn, Scott and Wolf, 2020.

[4]Riesinger *et al.*, 2020.

[5]Mädl *et al.*, 2016.

7.3 EXPERIMENTAL

The synthesis of Ni‴P_3 followed the literature procedure.[6] Orange crystals suitable for a diffraction experiment crystallised in $P2_1/c$ from a dichloromethane solution at room temperature. The Rigaku Oxford Diffraction SuperNova (Dualflex) Atlas S2 diffractometer irradiated the sample with Mo K_α radiation ($\lambda = 0.71073$ Å) and collected diffraction data. `Rigaku CrysAlis`$^{\text{Pro}}$ processed the obtained dataset.[7] The dataset is 99.94 % complete, with R_{int} of 2.93 %. Table C.1 brings further statistics on data collection and refinement.

The `Olex2` graphical user interface served to solve and refine the Ni‴P_3 crystal structure of Cp‴NiP_3.[8] It employed the external program ShelXT 2018/2 for the structure solution[9] and the internal program olex2.refine for structure refinement.[10] The last program removed symmetry-forbidden reflections from the `hkl` file and averaged the remaining ones.

The program `MoPro` refined the crystal structure using the Hansen-Coppens multipolar model.[11] A gradual refinement of following parameters was executed: the scaling factor coefficient, atomic coordinates, anisotropic displacement parameters, valence populations, multipolar coefficients (up to dipolar level for hydrogen atoms; octapolar levels for carbon and oxygen atoms; hexadecapolar level for the nickel and phosphorous atoms), expansion contraction parameters κ and κ' and anomalous dispersion (f' and f'') parameters for nickel.

A calculation on the SHADE3 server determined anisotropic displacement parameters of hydrogen atoms for the Cp‴ ligand.[12] Calculated values of U_{ani}(H) were constrained to those from SHADE3 during the refinement. In addition, several *R*-free calculations with 20 refinement cycles determined optimal restraint values for multipolar symmetry and valence/multipolar population similarities.

The model quality was evaluated through four graphical techniques: the DR plot (figure C.1),[13] the resolution plot (figure C.2),[14] the fractal dimension plot (figure C.3), and the residual density plots (figure C.4). `XDRKplots` made the first two plots, `jnk2RDA` created the third, and `VMoPro` plotted two-dimensional residual densities.[15] The grounds to decide in which directions to proceed with the refinement were the compromise between improving these plots, residual density global minimum/-maximum values and static deformation density contour plots.

`VMoPro` calculated total charge density 3D maps for the entire complex. `VMoPro` further calculated electrostatic potentials, bond critical points and their parameters (density, ellipticity, density Lapla-

[6] Ibid.
[7] Rigaku Oxford Diffraction, 2020.
[8] Dolomanov *et al.*, 2009.
[9] Bourhis *et al.*, 2015b.
[10] Bourhis *et al.*, 2015a.
[11] Jelsch *et al.*, 2005.
[12] Madsen, 2006; Madsen and Hoser, 2015.
[13] Abraham and Keve, 1971.
[14] Zhurov, Zhurova and Pinkerton, 2008.
[15] Meindl and Henn, 2008; Meindl, 2008; Jelsch *et al.*, 2005.

cian *etc.*) and generated two-dimensional property plots. `MoProViewer` created three-dimensional representations of molecules with charge density and ESP surfaces.[16]

7.4 Analysis

The first logical point of analysis for a charge density is the deformation density (figure 7.1). The cross-section of the nickel density perpendicular to the *cyclo*-P_3 plane shows an expected pattern of 3*d* orbital density (figure 7.1a); one can visualise this better on a three-dimensional representation (figure 7.2a). Orbital population analysis shows that most electrons reside in the xy plane (parallel to the *cyclo*-P_3 plane): two in the $d_{x^2-y^2}$ orbital, and another two in the d_{xy} orbital. The d_{z^2} and d_{xz} orbitals split further three electrons roughly in half, and the d_{yz} comes last with a mere one electron (table 7.1).

Strained, banana-like bonds bind together the *cyclo*-P_3 ring (figures 7.1b and 7.2b), a common feature seen in small rings.[17] The cyclopentadienyl moiety Cp‴ shows expected density patterns of an organic anion 7.1c), with the negative charge seemingly focused at the cyclopentadienyl ring.

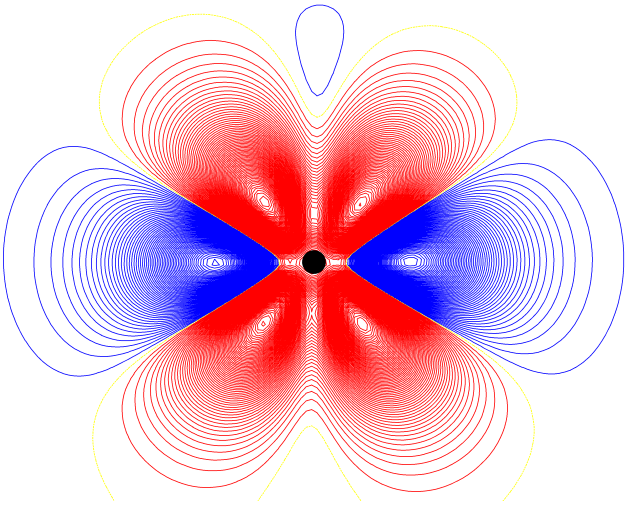
The charge analysis (table 7.2) concurs that the cyclopentadienyl ring holds a sizeable negative charge (a Bader charge of $-0.59\,e$, and a model $N_{\mathrm{val}} - P_{\mathrm{val}}$ charge of $-0.08\,e$). The electron-donating *tert*-butyl groups stabilise this charge. The model charge predicts that *tert*-butyl hydrogen atoms hold a positive, and carbon atoms a negative charge, but the Bader charge surprisingly contradicts this notion. The Cp‴ ESP plot at the $0.5\,e$ total density isosurface (figure 7.3a) favours the model charge interpretation. In both cases, however, the whole Cp‴ moiety carries a positive charge (Bader: $+0.14\,e$, model: $+0.80\,e$).

The *cyclo*-P_3 moiety carries the counter-charge (Bader: $-0.97\,e$, model: $-0.69\,e$) in the complex. The isolobality concept bolsters this finding: white phosphorous P_4 has strongly distorted lone pairs by some kind of aromaticity effect, and the negative charge on P_3 is consistent with that idea. In addition, the electronegativity difference between nickel and phosphorous atoms manifests itself in their charges.

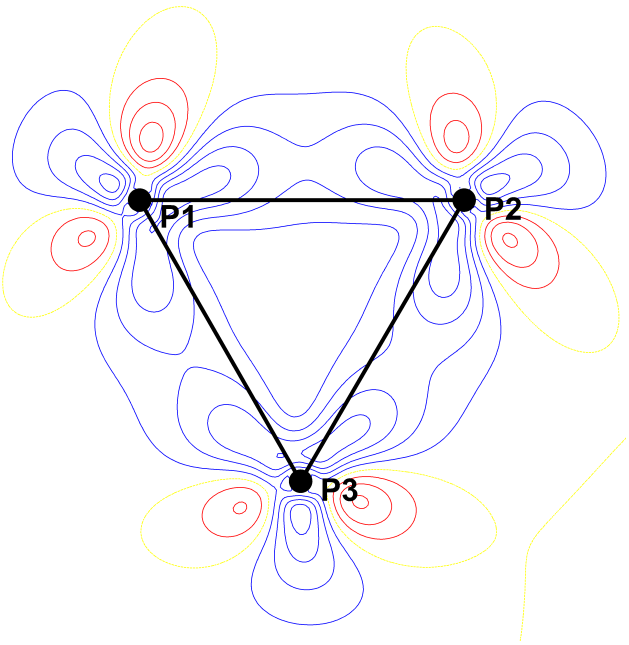
Table 7.1. Populations of nickel *d*-orbitals in Cp‴NiP_3.

Orbital	Population	Proportion (%)
P_{z^2}	1.65	20.2
P_{xz}	1.31	16.0
P_{yz}	1.14	13.9
$P_{x^2-y^2}$	2.00	24.4
P_{xy}	2.09	25.5
Total	8.18	100

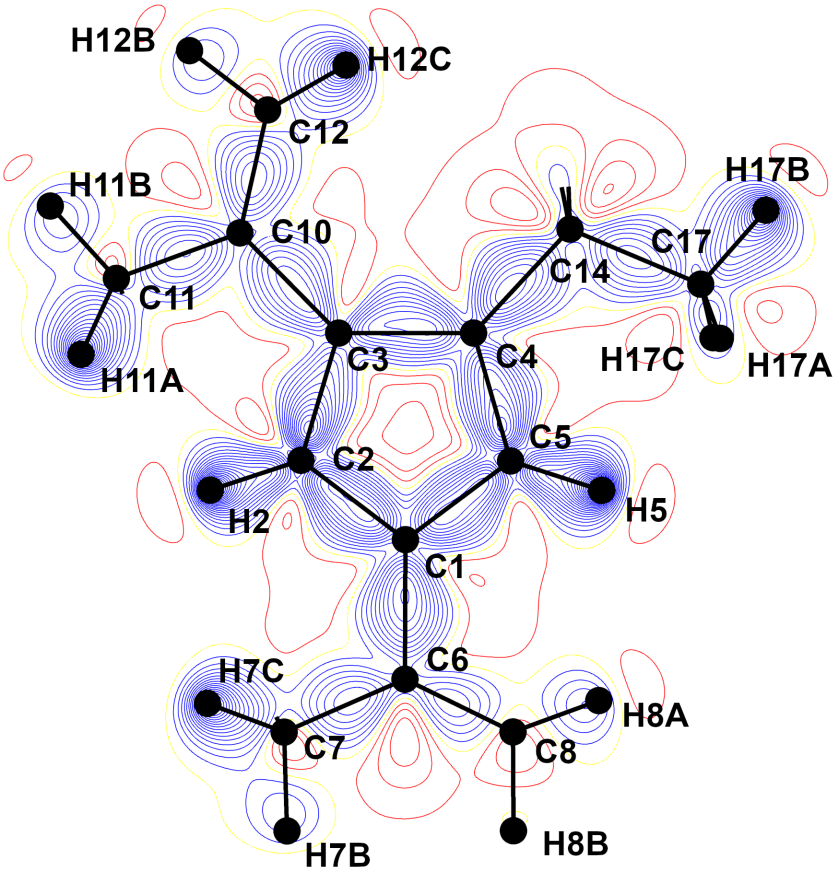
[16] Guillot *et al.*, 2014.
[17] Bak and Led, 1969.



(a) Ni.

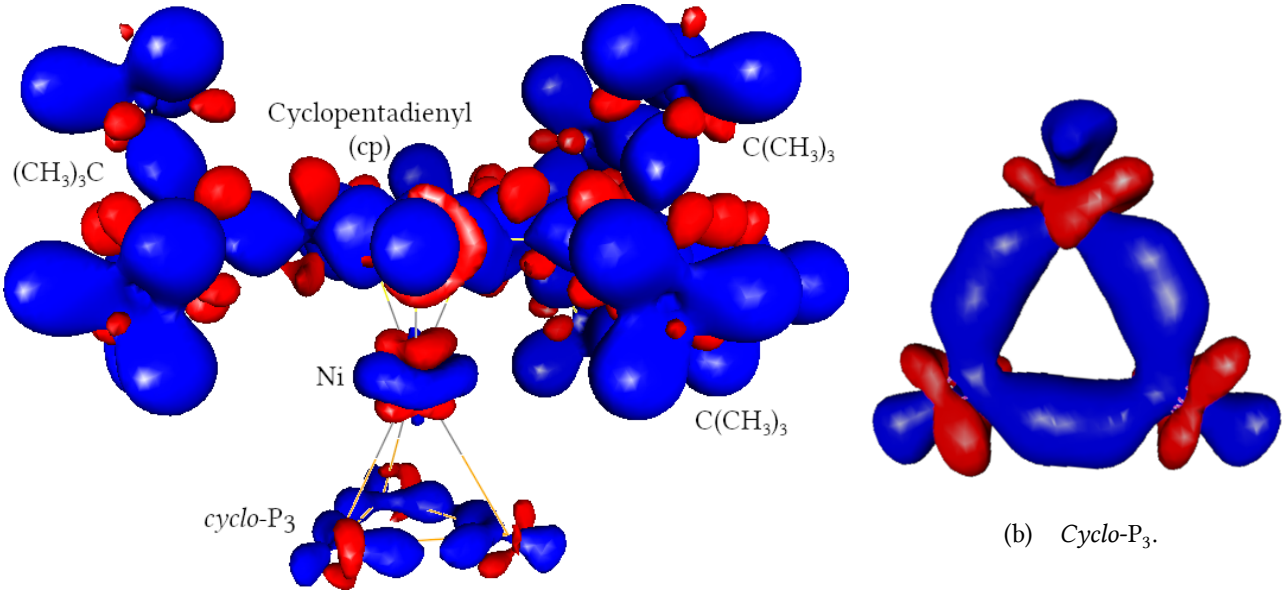


(b) *Cyclo*-P_3.



(c) Cp‴.

Figure 7.1. The two-dimensional static charge density deformation plots of various components of Cp‴NiP_3.



(a) Cp‴NiP_3.

(b) *Cyclo*-P_3.

Figure 7.2. The three-dimensional charge density deformation plots.

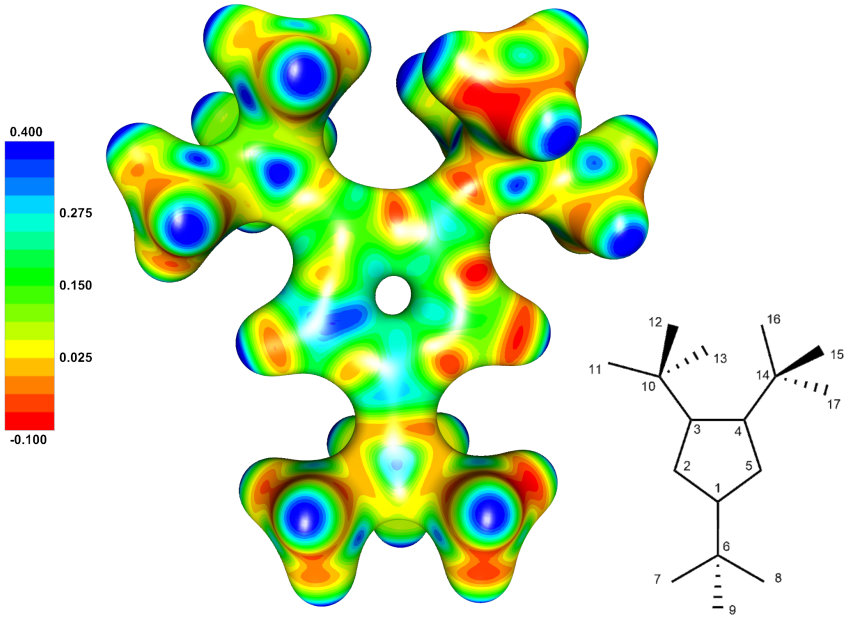
A strong negative ESP lies particularly in the P-P bonding region (figures 7.3b and 7.3c). The nickel atom is sandwiched between them, and its charge is $-0.18\,e$ according to the model ($N_{\text{val}} - P_{\text{val}}$). The Bader analysis shows a different picture, where the nickel has a significant positive charge ($+1.16\,e$), consistent with the ESP (figure 7.3d). This would suggest the existence of electrostatic tension between Ni and Cp‴.

`Charger` does not detect any such tension as what appears from model charges. In fact, the electrostatic interaction energy between Ni and Cp‴ (−274 kcal/mol) is on the same order of magnitude as that between Ni and *cyclo*-P_3 (−298 kcal/mol). Additionally, one can see the "fingerprint" of the cyclopentadienyl ring on the nickel ESP (figure 7.3e). The interaction between Cp‴ and *cyclo*-P_3 is expectedly smaller since they are further away (−41 kcal/mol). Overall, the Cp‴NiP_3 complex is fairly stable from the electrostatic point of view.

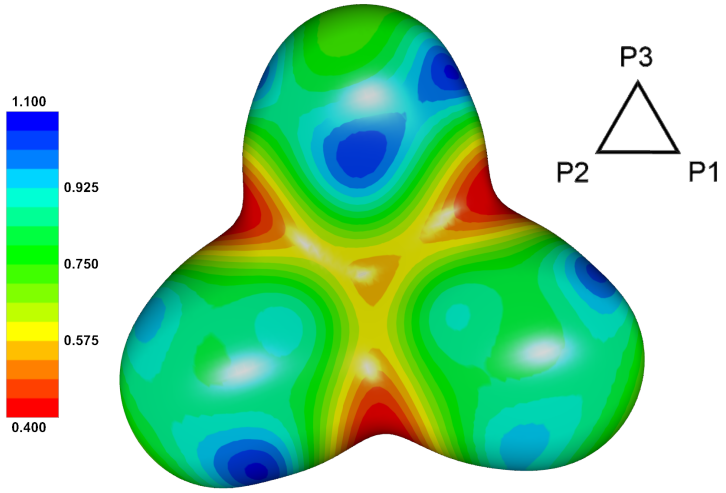
The intramolecular critical point analysis does not yield anything unexpected. The Ni-C and Ni-P contacts are similar in strength (ρ_{cp}) and largely coordinating (positive Laplacian values). The P-P contacts are stronger and not covalent (small positive Laplacian values). The covalent (negative Laplacian values) C-C and C-H bonds are typically even stronger.

The intermolecular contacts are surprisingly few. Only two C-H···P and a handful of C-H···C contacts have the interatomic distance shorter than the sum of their van der Waals radii. The critical points are weaker than those for typical hydrogen bonds, and the geometry does not seem to particularly favour the interaction.

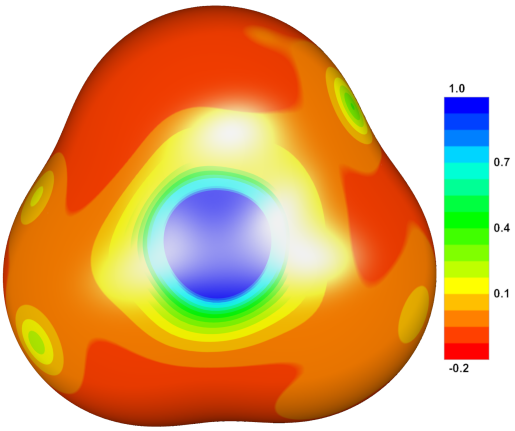
One could be forgiven for the temptation to see the Cp‴NiP_3 crystal as consisting of largely uncharged, roughly spheroidal entities with few contacts between them. This view proclaims that the crystal electrostatic interaction energy would not be significant. Before embarking to investigate it, a short detour into another fascinating compound is in order.



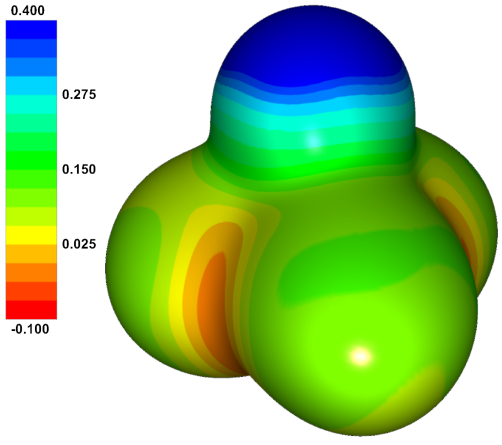
(a) 2,4,5-*Tris*(tertbutyl)cyclopentadienyl, $\rho = 0.5\,e\,\text{Å}^{-3}$.



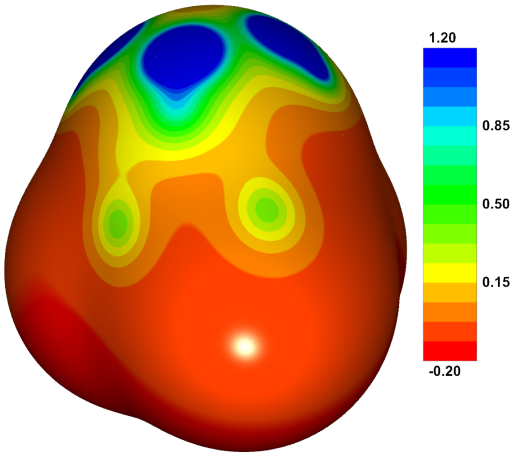
(b) *Cyclo*-P_3, $\rho = 0.5\,e\,\text{Å}^{-3}$.



(c) *Cyclo*-P_3, $\rho = 0.005\,e\,\text{Å}^{-3}$.



(d) Ni*Cyclo*-P_3, $\rho = 0.2\,e\,\text{Å}^{-3}$.



(e) Ni*Cyclo*-P_3, $\rho = 0.005\,e\,\text{Å}^{-3}$.

Figure 7.3. The electrostatic potential (scales in $e\,\text{Å}^{-1}$) of various components of Cp‴NiP_3 mapped onto different isodensity surfaces.

Table 7.2. Atomic charges calculated from the Cp‴NiP$_3$ multipolar model. Bader charges come from charge density integration of atomic basins, while model charges are $q_{\text{model}} = N_{\text{val}} - P_{\text{val}}$.

Atom	q_{Bader}/e	q_{model}/e	Atom	q_{Bader}/e	q_{model}/e
Ni1	+1.16	−0.18	C11	+0.05	−0.19
P1	−0.31	−0.23	H11A	−0.01	+0.10
P2	−0.36	−0.23	H11B	0.00	+0.10
P3	−0.30	−0.23	H11C	−0.01	+0.10
C1	−0.10	−0.15	C12	+0.08	−0.19
C2	−0.10	+0.07	H12A	−0.02	+0.10
H2	−0.03	+0.01	H12B	0.00	+0.10
C3	−0.08	+0.07	H12C	−0.01	+0.10
C4	−0.10	−0.15	C13	+0.06	−0.19
C5	−0.12	+0.07	H13A	0.00	+0.10
H5	−0.06	0.00	H13B	0.00	+0.10
C6	+0.06	−0.05	H13C	+0.01	+0.10
C7	+0.08	−0.18	C14	+0.07	−0.06
H7A	−0.02	+0.10	C15	+0.08	−0.18
H7B	+0.00	+0.10	H15A	−0.01	+0.10
H7C	+0.01	+0.10	H15B	+0.00	+0.10
C8	+0.04	−0.18	H15C	−0.01	+0.11
H8A	+0.02	+0.10	C16	+0.05	−0.18
H8B	+0.00	+0.10	H16A	+0.01	+0.10
H8C	+0.00	+0.10	H16B	+0.01	+0.10
C9	+0.05	−0.18	H16C	0.00	+0.10
H9A	+0.01	+0.10	C17	+0.03	−0.19
H9B	+0.01	+0.10	H17A	0.00	+0.10
H9C	+0.00	+0.10	H17B	+0.01	+0.10
C10	+0.08	−0.06	H17C	0.00	+0.10
Ni	+1.16	−0.18	Cp‴	+0.14	+0.80
cyclo-P$_3$	−0.97	−0.69	Total	+0.33	−0.07

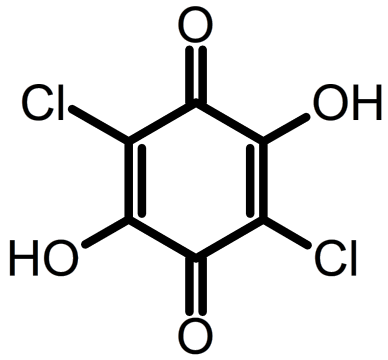
The real total model charge is 0.00 e, but the rounding resulted in the −0.07 e charge.

7.5 Chloranilic Acid

This section reflects on author's first charge density love: the chloranilic acid (H_2CA).[18] It is a quinoid similar to the dihydroxyquinonate anion described in the previous chapter, except with chlorine substituents on C3 carbon atoms (scheme 7.1). It dissolves into water, forming a deep purple solution, from which orange single crystals of chloranilic acid dihydrate ($H_2CA \cdot 2\,H_2O$) grow fairly quickly. Those crystals diffract the X-ray beam very well, and one can model a high-resolution structure almost effortlessly.

The experimental procedure mirrors the one for **LiDHQ**, except that the Rigaku Oxford Diffraction Supernova diffractometer was used for data collection, and CrysAlis$^{\text{Pro}}$ for dataset processing. The

[18] Vuković, Molčanov *et al.*, 2019.



Scheme 7.1. The chloranilic acid (H_2CA).

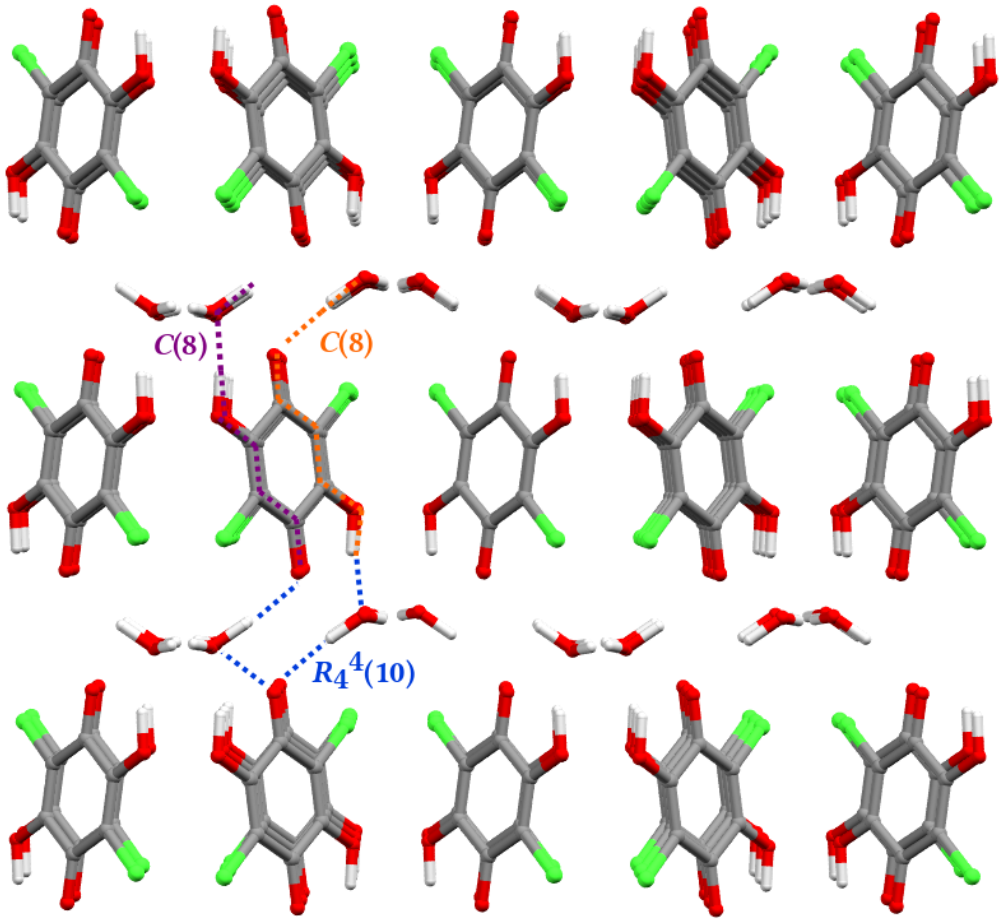


Figure 7.4. The chloranilic acid packing in the crystal *c* direction. The chains [$C(8)$] and rings [$R_4^4(10)$] of hydrogen bonds are depicted explicitly. The π-stacks of chloranilic acid moieties are parallel with the viewing direction.

overall quality is on the same level as in the **LiDHQ** case, largely due to chlorine atoms that absorb the X-ray beam and whose charge density disturbs the Hansen-Coppens model. Vuković, Molčanov *et al.* detail refinement, modelling, and model exploration.[19]

[19] Ibid.

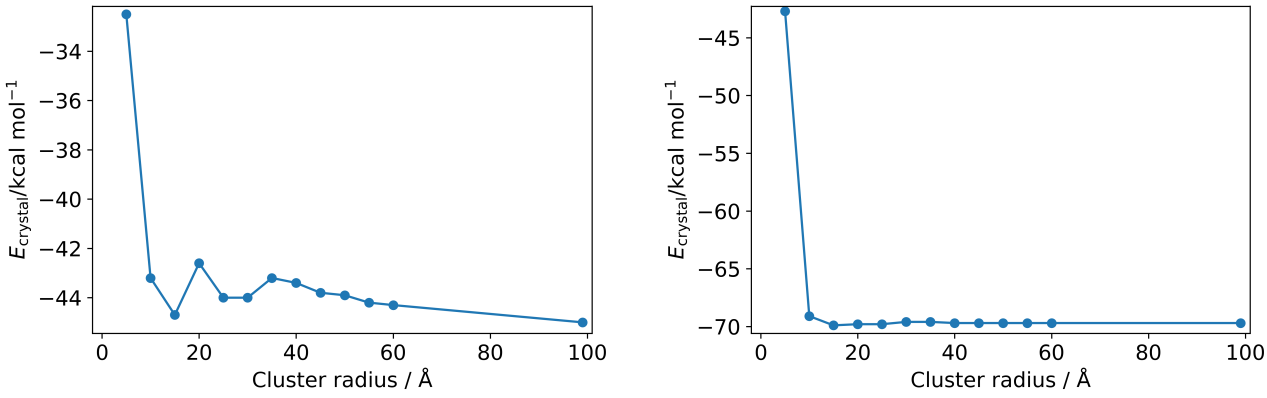
The major interactions in the crystal packing of $H_2CA \cdot 2\,H_2O$ are the hydrogen bonds between water and chloranilic acid moieties. Furthermore, cycle stacking also plays a large role (figure 7.4, electrostatic interaction energy of −5.55 kcal/mol). These strong electrostatic contributions put the chloranilic acid dihydrate forward as a good model for crystal electrostatic energy estimation, along with Cp‴NiP_3. In addition, these two compounds give `Charger` three more atoms to test: phosphorous, chlorine and nickel.

7.6 Crystal Energy

Now that the two main actors (H_2CA and Cp‴NiP_3) are well known, the focus can finally turn towards estimating electrostatic interaction energy of a crystal lattice. One estimates this energy by taking a cluster of molecules and calculating the electrostatic interaction energy of the central molecule (denoted by the ORTEP code 55501) and all the other molecules. The larger the cluster, the better the estimate, and one could in principle use Fourier-based techniques to estimate the energy between the central molecule and all other pairs in the infinite crystal lattice.

`MoProViewer` can generate molecular clusters of arbitrary size around a central molecule, such that the centres of geometry of all generated molecules fall within a certain radius from the central molecule's geometry centre. It has an internal limit: it relies on ORTEP codes to keep track of generated molecules. It can thus not generate molecules beyond 5 unit cells away in each direction from the central molecule. This makes larger clusters not fully spherical.

One idea was to check how large a cluster one needs to get a reliable estimate of electrostatic interaction energy. The answer is surprisingly low: even a 10 Å cluster can provide a good assessment (figure 7.5). The larger Cp‴NiP_3 took a larger cluster radius to converge (figure 7.5a) than the smaller



(a) Cp‴NiP_3.

(b) Chloranilic acid dihydrate ($H_2CA \cdot 2\,H_2O$).

Figure 7.5. Electrostatic interaction energy of crystal packing ($E_{crystal}$) as a function of cluster size. `Charger` determines $E_{crystal}$ as the interaction energy between the central (ORTEP code 55501) molecule and the surrounding molecules (the "cluster") in the sphere of arbitrary radius. In the case of chloranilic acid dihydrate, the central "molecule" consists of three molecules: one chloranilic acid molecule and two water molecules.

H_2CA (figure 7.5b). The ORTEP limit makes the H_2CA clusters flatter in one dimension, which might contribute to this observation. The energies between components of the asymmetric units have not been considered in this calculation.

The $Cp'''NiP_3$ electrostatic crystal energy evaluates to around a non-negligible −44 kcal/mol. This is in stark contrast with the idea that one should view $Cp'''NiP_3$ as a quasi-neutral spheroid. Trying to ascertain which contacts are the most responsible for this energy might prove a fruitful endeavour.

The converged $H_2CA \cdot 2\,H_2O$ electrostatic crystal energy amounts to −69.7 kcal/mol, roughly 13 times more than the offset cycle stacking (−5.55 kcal/mol). Only two H_2CA moieties stack up next to the central molecule, so the main contributors to H_2CA electrostatic stability must be coming from hydrogen bonds.

Part III

Postface

8 Conclusions

8.1 Français

8.1.1 La méthode aEP/pMM

La méthode du potentiel exact analytique / moments pseudo-multipolaires (PEa/MpM – ou, selon l'acronyme anglais, aEP/pMM) repose en grande partie sur la fonction α de Löwdin, et plus particulièrement sur le développement de Nguyen, Kisiel et Volkov (2018). Le traitement mathématique explicite de l'intégrale de Coulomb pour une densité de type Slater ouvre les portes à un calcul plus rapide, plus robuste et plus fiable de l'énergie d'interaction électrostatique. Les fonctions radiales de type Slater sont au cœur de la modélisation expérimentale de la densité de charge (le modèle Hansen-Coppens) : ce développement constitue un autre outil pour quantifier et examiner les propriétés observées à l'échelle microscopique.

Les moments pseudo-multipolaires de Buckingham sont une approximation valable à longue distance. Le code existant de la partie aEP a pu être réutilisé pour construire la partie pMM, mais elle ne met pas entièrement en œuvre les moments multipolaires de Buckingham. La faiblesse de pMM réside dans l'intégration complète de certaines interactions multipolaires : il est donc plus lent qu'un calcul de moments multipolaires de Buckingham ponctuels ordinaire. Néanmoins, cette approche unique pourrait être adaptée, où une mise en œuvre correcte des moments multipolaires pourrait être entreprise, afin d'améliorer la vitesse calcul autant que possible, en particulier pour les grands systèmes.

8.1.2 `Charger`

`Charger` est une implémentation indépendante de la méthode aEP/pMM pour calculer l'énergie d'interaction électrostatique. Il sert à réaliser ce calcul de façon beaucoup plus rapide. La librairie `Charger` est entièrement implémentée en `C` et disponible sous la licence permissive Apache 2.0. Le code source est disponible en téléchargement ici :

`https://crm2.univ-lorraine.fr/lab/fr/software/mopro/`

Le programme de visualisation moléculaire `MoProViewer` contient une interface pour `Charger`. Cette interface facilite l'accès rapide aux énergies d'interaction électrostatique entre diverses molécules sans devoir comprendre la logique interne de `Charger`. L'auteur espère que cette facilité d'utilisation du calcul des énergies d'interaction électrostatique grâce à l'interface graphique encouragera leur utilisation à grande échelle dans les communautés de la biologie structurale et de la chimie quantique. Le programme `MoProViewer` avec `Charger` se trouve sur le même hyperlien.

8.1.3 Applications

Charger a été appliqué avec succès à un large éventail de systèmes. Il est adapté aux calculs rapides sur de grands systèmes comme les protéines, où il calcule l'énergie protéine-ligand en quelques secondes. Il a également traité des systèmes de petites molécules contenant des atomes d'hydrogène, de lithium, de carbone, d'azote, d'oxygène, de phosphore, de soufre, de chlore et de nickel. Il a été testé contre les charges ponctuelles AMBER03, avec les résultats prometteurs.

Les énergies d'interaction électrostatique plus élevées sont en corrélation avec les décalages thermiques expérimentaux obtenus sur les systèmes protéine-ligand. Elles permettent également d'identifier les acides aminés qui contribuent de manière significative à la liaison du ligand, et ceux qui l'entravent. Une investigation détaillée du champ électrique dans le site actif confirment ce pouvoir prédictif. De plus, des résultats de `Charger` peuvent indiquer des candidats utiles pour une étude de mutation.

`Charger` complète l'analyse simple, basée sur les charges, des contacts proches. Il confirme souvent les conclusions obtenues par une analyse simple (par ex. le contact DMSO···DHQ^{2-}). Il est parfois surprenant (par ex. dans l'investigation de l'interaction entre le Cp''' et Ni, les deux étant chargés positivement), en offrant une perspective plus nuancée puisqu'il prend en compte l'ensemble de la densité de charge, plutôt que de se limiter sur les charges ponctuelles.

`Charger` fournit également un moyen simple de déduire l'énergie d'interaction électrostatique d'un cristal. Cette énergie peut être étonnamment forte, même en l'absence de contacts électrostatiques évidents entre les molécules qui composent le cristal.

8.2 English

8.2.1 The aEP/pMM Method

The analytical exact potential / pseudo-multipolar moments (aEP/pMM) method rests largely on the Löwdin's α function, and specifically on the development by Nguyen, Kisiel and Volkov (2018). The explicit mathematical treatment of the Coulomb integral for a Slater-type density opens the doors to a faster, more robust and more reliable calculation of electrostatic interaction energy. Slater-type radial functions rest at the core of experimental charge density modelling (the Hansen-Coppens model): this development creates another tool to quantify and examine observed properties at the microscopic level.

The approximation of Buckingham pseudo-multipolar moments is an approximation valid at long distance. The existing code from the aEP part was re-used to build the pMM part, but it does not entirely implement those multipolar moments. The weakness of pMM lies in full integration of some multipole interactions: it is thus slower than an ordinary multipolar moments calculation. Nevertheless, this unique approach could be further adapted, or a proper multipolar moments implementation undertaken, to improve the calculation speed as much as possible, particularly for large systems.

8.2.2 Charger

Charger is an independent implementation of the aEP/pMM method to calculate electrostatic interaction energy. It serves to calculate the electrostatic energy in a much faster way. The Charger library is fully implemented in C and available under the permissive Apache 2.0 licence. The source code is available for download here:

https://crm2.univ-lorraine.fr/lab/fr/software/mopro/

The molecular visualisation program MoProViewer contains a binding to Charger. This facilitates a fast access to electrostatic interaction energies between various molecules without the need to understand the internal mechanics of Charger. The author hopes that this low barrier to entry to electrostatic interaction energies shall encourage their wide use in the structural biology and quantum chemistry communities. The MoProViewer with Charger resides on the same link.

8.2.3 Applications

Charger has been successfully applied to a wide range of systems. It is suitable for fast calculations on large systems like proteins, where it calculates the protein-ligand energy in a matter of seconds. It also dealt with small molecule systems containing hydrogen, lithium, carbon, nitrogen, oxygen, phosphorous, sulphur, chlorine and nickel atoms. It was further tested against AMBER03 point charges, with encouraging results.

Charger electrostatic interaction energies correlate with experimental thermal shifts obtained on a protein-ligand system. They also pinpoint which amino acids contribute significantly to ligand binding, and which hinder it. More detailed analyses of the electric field in the binding site vindicate this predictive power. Furthermore, Charger's results can point to useful mutation candidates for a mutation study.

Charger complements simple, charge-based analysis of close contacts. It often underscores such conclusions (*e.g.* the DMSO···DHQ^{2-} close contact). It sometimes surprises (*e.g.* when investigating the interaction between positively charged Cp‴ and positively charged Ni), providing a more nuanced perspective since it takes the entire charge density into account, rather than point charges only.

Charger also provides a simple way to deduce electrostatic interaction energy from a crystal. This energy can be surprisingly strong even when no obvious electrostatic contacts occur between the moieties that reside within it.

Bibliography

General Introduction

Bojarowski, S. A., Kumar, P. and Dominiak, P. M. (2016). A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *ChemPhysChem* **17**.16, 2455–2460. DOI: 10.1002/cphc.201600390.

Buckingham, A. D. (1959). Molecular quadrupole moments. *Quarterly Reviews, Chemical Society* **13**.3, 183. DOI: 10.1039/qr9591300183.

Cochran, W. (1956). A comparison of calculated and measured electron distributions in the benzene ring. *Acta Crystallographica* **9**.11, 924–928. DOI: 10.1107/s0365110x5600259x.

Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. and Yang, Y. W. (1979). Net atomic charges and molecular dipole moments from spherical-atom X-ray refinements, and the relation between atomic charge and shape. *Acta Crystallographica Section A* **35**.1, 63–72. DOI: 10.1107/s0567739479000127.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. International Union of Crystallography, Oxford University Press, Oxford.

Coulomb, C.-A. de (1785). Second mémoire sur l'électricité et le magnétisme. *Histoire de l'Académie Royale des Sciences*, 612–638.

Dittrich, B., Koritsánszky, T. and Luger, P. (2004). A Simple Approach to Nonspherical Electron Densities by Using Invarioms. *Angewandte Chemie International Edition* **43**.20, 2718–2721. DOI: 10.1002/anie.200353596.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. and Jelsch, C. (2012). An improved experimental databank of transferable multipolar atom models – ELMAM2. Construction details and applications. *Acta Crystallographica Section A Foundations of Crystallography* **68**.3, 337–351. DOI: 10.1107/s0108767312008197.

Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. and Coppens, P. (2006). A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *Journal of Chemical Theory and Computation* **3**.1, 232–247. DOI: 10.1021/ct6001994.

Hansen, N. K. and Coppens, P. (1978). Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallographica Section A* **34**.6, 909–921. DOI: 10.1107/s0567739478001886.

Hirshfeld, F. L. (1971). Difference densities by least-squares refinement: fumaramic acid. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* **27**.4, 769–781. DOI: 10.1107/s0567740871002905.

Huang, L., Massa, L. and Karle, J. (2005). Kernel Energy Method: Application to DNA. *Biochemistry* **44**.50, 16747–16752. DOI: 10.1021/bi0516551.

Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods usingMoPro. *Journal of Applied Crystallography* **38**.1, 38–54. DOI: 10.1107/s0021889804025518.

Kumar, P., Gruza, B., Bojarowski, S. A. and Dominiak, P. M. (2019). Extension of the transferable aspherical pseudoatom data bank for the comparison of molecular electrostatic potentials in structure–activity studies. *Acta Crystallographica Section A Foundations and Advances* **75**.2, 398–408. DOI: 10.1107/s2053273319000482.

Leach, A. (2001). *Molecular modelling : principles and applications*. Harlow, England New York: Prentice Hall.

Lennard-Jones, J. E. (1924a). On the determination of molecular fields. – I. From the variation of the viscosity of a gas with temperature. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106**.738, 441–462. DOI: 10.1098/rspa.1924.0081.

— (1924b). On the determination of molecular fields. – II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106**.738, 463–477. DOI: 10.1098/rspa.1924.0082.

— (1931). Cohesion. *Proceedings of the Physical Society* **43**.5, 461–482. DOI: 10.1088/0959-5309/43/5/301.

Mandal, S. K., Saha, P., Munshi, P. and Sukumar, N. (2017). Exploring potent ligand for proteins: insights from knowledge-based scoring functions and molecular interaction energies. *Structural Chemistry* **28**.5, 1537–1552. DOI: 10.1007/s11224-017-1007-y.

Massa, L., Keith, T., Cheng, Y. and Matta, C. F. (2019). The kernel energy method applied to quantum theory of atoms in molecules – energies of interacting quantum atoms. *Chemical Physics Letters* **734**, 136650. DOI: 10.1016/j.cplett.2019.136650.

Newton, I. S. (1686). *Philosophiæ Naturalis Principia Mathematica*. Societas Regiæ, Londinus.

Nguyen, D., Kisiel, Z. and Volkov, A. (2018). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. I. The Löwdin α-function method. *Acta Crystallographica Section A Foundations and Advances* **74**.5, 524–536. DOI: 10.1107/s2053273318008690.

Nguyen, D., Macchi, P. and Volkov, A. (2020). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. III. Application to crystal structures via the Ewald and direct summation methods. *Acta Crystallographica Section A Foundations and Advances* **76**.6, 630–651. DOI: 10.1107/s2053273320009584.

Nguyen, D. and Volkov, A. (2019). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. II. The Fourier transform method. *Acta Crystallographica Section A Foundations and Advances* **75**.3, 448–464. DOI: 10.1107/s2053273319002535.

Tomiie, Y. (1958). The Electron Distribution and the Location of the Bonded Hydrogen Atom in Crystals. *Journal of the Physical Society of Japan* **13**.9, 1030–1037. DOI: 10.1143/jpsj.13.1030.

Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. and Koritsanszky, T. (2016). *XD2016 - A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors*.

Volkov, A., King, H. F., Coppens, P. and Farrugia, L. J. (2006). On the calculation of the electrostatic potential, electric field and electric field gradient from the aspherical pseudoatom model.

Acta Crystallographica Section A Foundations of Crystallography **62**.5, 400–408. DOI: 10.1107/s0108767306026298.

Volkov, A., Koritsanszky, T. and Coppens, P. (2004). Combination of the exact potential and multipole methods (EP/MM) for evaluation of intermolecular electrostatic interaction energies with pseudoatom representation of molecular electron densities. *Chemical Physics Letters* **391**.1-3, 170–175. DOI: 10.1016/j.cplett.2004.04.097.

To Calculate Electrostatic Energy

Buckingham, A. D. (1959). Molecular quadrupole moments. *Quarterly Reviews, Chemical Society* **13**.3, 183. DOI: 10.1039/qr9591300183.

Clementi, E. and Roetti, C. (1974). Roothaan-Hartree-Fock atomic wavefunctions. *Atomic Data and Nuclear Data Tables* **14**.3-4, 177–478. DOI: 10.1016/s0092-640x(74)80016-1.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. International Union of Crystallography, Oxford University Press, Oxford.

Coulomb, C.-A. de (1785). Second mémoire sur l'électricité et le magnétisme. *Histoire de l'Académie Royale des Sciences*, 612–638.

Fournier, B., Bendeif, E.-E., Guillot, B., Podjarny, A., Lecomte, C. and Jelsch, C. (2009). Charge Density and Electrostatic Interactions of Fidarestat, an Inhibitor of Human Aldose Reductase. *Journal of the American Chemical Society* **131**.31, 10929–10941. DOI: 10.1021/ja8095015.

Griffiths, D. J. (2021). *Introduction to Electrodynamics*.

Hans-Jurgen Weber, G., Weber, H., Arfken, G., Holland, B., Library, E., Harris, F. and Singer, T. (2004). *Essential Mathematical Methods for Physicists, ISE*. Elsevier Science.

Hansen, N. K. and Coppens, P. (1978). Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallographica Section A* **34**.6, 909–921. DOI: 10.1107/s0567739478001886.

Johnson, B. G., Gill, P. M., Pople, J. A. and Fox, D. J. (1993). Computing molecular electrostatic potentials with the PRISM algorithm. *Chemical Physics Letters* **206**.1-4, 239–246. DOI: 10.1016/0009-2614(93)85547-2.

Jones, H. W. (1991). Analytic Löwdin alpha-function method for two-center electron-repulsion integrals over slater-type orbitals. *Journal of Computational Chemistry* **12**.10, 1217–1222. DOI: 10.1002/jcc.540121008.

— (1992). Semianalytical method for four-center molecular integrals over Slater-type orbitals. *International Journal of Quantum Chemistry* **42**.4, 779–784. DOI: 10.1002/qua.560420417.

— (1993). Benchmark values for two-center Coulomb integrals over slater-type orbitals. *International Journal of Quantum Chemistry* **45**.1, 21–30. DOI: 10.1002/qua.560450105.

Jones, H. W. and Weatherford, C. A. (1978). A modified form of Sharma's formula for STO Löwdin alpha functions with recurrence relations for the coefficient matrix. *International Journal of Quantum Chemistry* **14**.S12, 483–488. DOI: 10.1002/qua.560140842.

Jones, H. W. and Weatherford, C. A. (1989). The Löwdin α function and its application to the multicenter molecular integral problem over slater-type orbitals. *Journal of Molecular Structure: THEOCHEM* **199**, 233–243. DOI: 10.1016/0166-1280(89)80055-7.

Koga, T., Kanayama, K., Watanabe, T., Imai, T. and Thakkar, A. J. (2000). Analytical Hartree-Fock wave functions for the atoms Cs to Lr. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **104**.5, 411–413. DOI: 10.1007/s002140000150.

Koga, T., Tatewaki, H. and Thakkar, A. J. (1993). Roothaan-Hartree-Fock wave functions for atoms withZ≤54. *Physical Review A* **47**.5, 4510–4512. DOI: 10.1103/physreva.47.4510.

Koga, T., Watanabe, S., Kanayama, K., Yasuda, R. and Thakkar, A. J. (1995). Improved Roothaan-Hartree-Fock wave functions for atoms and ions with N≤54. *The Journal of Chemical Physics* **103**.8, 3000–3005. DOI: 10.1063/1.470488.

Löwdin, P.-O. (1956). Quantum theory of cohesive properties of solids. *Advances in Physics* **5**.17, 1–171. DOI: 10.1080/00018735600101155.

Michael, J. R. and Volkov, A. (2015). Density- and wavefunction-normalized Cartesian spherical harmonics forl≤ 20. *Acta Crystallographica Section A Foundations and Advances* **71**.2, 245–249. DOI: 10.1107/s2053273314024838.

Nguyen, D., Kisiel, Z. and Volkov, A. (2018). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. I. The Löwdin α-function method. *Acta Crystallographica Section A Foundations and Advances* **74**.5, 524–536. DOI: 10.1107/s2053273318008690.

Prokopi, N. (2021). *Calculation of Euler angles*. URL: http://geom3d.com/data/documents/Calculation=20of=20Euler=20angles.pdf (visited on 07/06/2021).

Slater, J. C. (1930). Atomic Shielding Constants. *Physical Review* **36**.1, 57–64. DOI: 10.1103/physrev.36.57.

Spackman, M. A. (2007). Comment on On the calculation of the electrostatic potential, electric field and electric field gradient from the aspherical pseudoatom modelby Volkov, King, Coppens & Farrugia (2006). *Acta Crystallographica Section A Foundations of Crystallography* **63**.2, 198–200. DOI: 10.1107/s0108767307001298.

Volkov, A. and Coppens, P. (2007). Response to Spackman's comment on *On the calculation of the electrostatic potential, electric field and electric field gradient from the aspherical pseudoatom model*. *Acta Crystallographica Section A Foundations of Crystallography* **63**.2, 201–203. DOI: 10.1107/s0108767307002620.

Volkov, A., King, H. F., Coppens, P. and Farrugia, L. J. (2006). On the calculation of the electrostatic potential, electric field and electric field gradient from the aspherical pseudoatom model. *Acta Crystallographica Section A Foundations of Crystallography* **62**.5, 400–408. DOI: 10.1107/s0108767306026298.

Volkov, A., Koritsanszky, T. and Coppens, P. (2004). Combination of the exact potential and multipole methods (EP/MM) for evaluation of intermolecular electrostatic interaction energies with pseudoatom representation of molecular electron densities. *Chemical Physics Letters* **391**.1-3, 170–175. DOI: 10.1016/j.cplett.2004.04.097.

Wahl, A. C., Cade, P. E. and Roothaan, C. C. J. (1964). Study of Two-Center Integrals Useful in Calculations on Molecular Structure. V. General Methods for Diatomic Integrals Applicable to Digital Computers. *The Journal of Chemical Physics* **41**.9, 2578–2599. DOI: 10.1063/1.1726326.

Charger

Abergel, R. and Moisan, L. (2020). Algorithm 1006: Fast and Accurate Evaluation of a Generalized Incomplete Gamma Function. *ACM Transactions on Mathematical Software* **46**.1, 1–24. DOI: 10.1145/3365983.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. and al., et (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**.16, 1999–2012. DOI: 10.1002/jcc.10349.

Fournier, B., Bendeif, E.-E., Guillot, B., Podjarny, A., Lecomte, C. and Jelsch, C. (2009). Charge Density and Electrostatic Interactions of Fidarestat, an Inhibitor of Human Aldose Reductase. *Journal of the American Chemical Society* **131**.31, 10929–10941. DOI: 10.1021/ja8095015.

Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods usingMoPro. *Journal of Applied Crystallography* **38**.1, 38–54. DOI: 10.1107/s0021889804025518.

Kumar, P., Bojarowski, S. A., Jarzembska, K. N., Domagała, S., Vanommeslaeghe, K., MacKerell, A. D. and Dominiak, P. M. (2014). A Comparative Study of Transferable Aspherical Pseudoatom Databank and Classical Force Fields for Predicting Electrostatic Interactions in Molecular Dimers. *Journal of Chemical Theory and Computation* **10**.4, 1652–1664. DOI: 10.1021/ct4011129.

Nguyen, D., Kisiel, Z. and Volkov, A. (2018). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. I. The Löwdin α-function method. *Acta Crystallographica Section A Foundations and Advances* **74**.5, 524–536. DOI: 10.1107/s2053273318008690.

Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. and Koritsanszky, T. (2016). *XD2016 - A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors*.

Volkov, A., Koritsanszky, T. and Coppens, P. (2004). Combination of the exact potential and multipole methods (EP/MM) for evaluation of intermolecular electrostatic interaction energies with pseudoatom representation of molecular electron densities. *Chemical Physics Letters* **391**.1-3, 170–175. DOI: 10.1016/j.cplett.2004.04.097.

Winkler, J. R. (1993). Numerical recipes in C: The art of scientific computing, second edition. *Endeavour* **17**.4, 201. DOI: 10.1016/0160-9327(93)90069-f.

Glutathione Transferase and Its Ligands

Cimmperman, P., Baranauskienė, L., Jachimovičiūtė, S., Jachno, J., Torresan, J., Michailovienė, V., Matulienė, J., Sereikaitė, J., Bumelis, V. and Matulis, D. (2008). A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands. *Biophysical Journal* **95**.7, 3222–3231. DOI: 10.1529/biophysj.108.134973.

Cox, P. J., Kechagias, D. and Kelly, O. (2008). Conformations of substituted benzophenones. *Acta Crystallographica Section B Structural Science* **64**.2, 206–216. DOI: 10.1107/s0108768108000232.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. and Jelsch, C. (2012). An improved experimental databank of transferable multipolar atom models – ELMAM2. Construction details and ap-

plications. *Acta Crystallographica Section A Foundations of Crystallography* **68**.3, 337–351. DOI: 10.1107/s0108767312008197.

Guillot, B., Enrique, E., Huder, L. and Jelsch, C. (2014). MoProViewer: a tool to study proteins from a charge density science perspective. *Acta Crystallographica Section A Foundations and Advances* **70**.a1, C279–C279. DOI: 10.1107/s2053273314097204.

Hooft, R. W., Sander, C. and Vriend, G. (1996). Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Structure, Function, and Genetics* **26**.4, 363–376. DOI: 10.1002/(sici)1097-0134(199612)26:4<363::aid-prot1>3.0.co;2-d.

Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods usingMoPro. *Journal of Applied Crystallography* **38**.1, 38–54. DOI: 10.1107/s0021889804025518.

Leduc, T. (2019). "Vers un potentiel multipolaire quantitatif et transférable aux macromolécules biologiques : une étude méthodologique des effets de la polarisabilité." PhD Thesis. Université de Lorraine.

Leduc, T., Aubert, E., Espinosa, E., Jelsch, C., Iordache, C. and Guillot, B. (2019). Polarization of Electron Density Databases of Transferable Multipolar Atoms. *The Journal of Physical Chemistry A* **123**.32, 7156–7170. DOI: 10.1021/acs.jpca.9b05051.

Levine, I. (2009). *Physical Chemistry*. 6th ed.

Perrot, T., Schwartz, M., Deroy, A., Girardet, J.-M., Kohler, A., Morel-Rouhier, M., Favier, F., Gelhaye, E. and Didierjean, C. (2021). Diversity of Omega Glutathione Transferases in mushroom-forming fungi revealed by phylogenetic, transcriptomic, biochemical and structural approaches. *Fungal Genetics and Biology* **148**, 103506. DOI: 10.1016/j.fgb.2020.103506.

Perrot, T., Schwartz, M., Saiag, F., Salzet, G., Dumarçay, S., Favier, F., Gérardin, P., Girardet, J.-M., Sormani, R., Morel-Rouhier, M. and al., et (2018). Fungal Glutathione Transferases as Tools to Explore the Chemical Diversity of Amazonian Wood Extractives. *ACS Sustainable Chemistry & Engineering* **6**.10, 13078–13085. DOI: 10.1021/acssuschemeng.8b02636.

Schrödinger, LLC and DeLano, W. (20, 2020). *PyMOL*. Version 2.4.0.

Schwartz, M., Perrot, T., Aubert, E., Dumarçay, S., Favier, F., Gérardin, P., Morel-Rouhier, M., Mulliert, G., Saiag, F., Didierjean, C. and al., et (2018). Molecular recognition of wood polyphenols by phase II detoxification enzymes of the white rot Trametes versicolor. *Scientific Reports* **8**.1. DOI: 10.1038/s41598-018-26601-3.

Sheehan, D., Meade, G., Foley, V. M. and Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal* **360**.1, 1–16. DOI: 10.1042/bj3600001.

Thapa, B. and Raghavachari, K. (2019). Energy Decomposition Analysis of Protein–Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method. *Journal of Chemical Information and Modeling* **59**.8, 3474–3484. DOI: 10.1021/acs.jcim.9b00432.

LiDHQ

Abraham, S. and Keve, E. T. (1971). Normal probability plot analysis of error in measured and derived quantities and standard deviations. *Acta Cryst.* **A27**, 157–165.

Angarov, V. and Kozuch, S. (2018). On the σ, π and δ hole interactions: a molecular orbital overview. *New Journal of Chemistry* **42**.2, 1413–1422. DOI: 10.1039/c7nj03632a.

Bader, R. F. W. (2009). Bond Paths Are Not Chemical Bonds. *The Journal of Physical Chemistry A* **113**.38, 10391–10396. DOI: 10.1021/jp906341r.

Bartashevich, E., Nikulov, D., Vener, M. and Tsirelson, V. (2011). QTAIMC study of the X–H/H···O bond order indices (X=O, N, C) in molecular systems. *Comput. Theor. Chem.* **973**.1-3, 33–39. DOI: 10.1016/j.comptc.2011.06.025.

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**.7, 5648–5652. DOI: 10.1063/1.464913.

Blessing, R. H. (1995). An empirical correction for absorption anisotropy. *Acta Crystallographica Section A Foundations of Crystallography* **51**.1, 33–38. DOI: 10.1107/s0108767394005726.

Bourhis, L., Dolomanov, O., Gildea, R., Howard, J. and Puschmann, H. (2015b). *ShelXT*.

Bruker (2012). *SAINT*. Bruker AXS Inc., Madison, Wisconsin, SAD.

Clark, T., Chandrasekhar, J., Spitznagel, G. W. and Schleyer, P. V. R. S. (1983). Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *Journal of Computational Chemistry* **4**.3, 294–301. DOI: 10.1002/jcc.540040303.

Dolomanov, O. V., Bourhis, L., Gildea, R. and Puschmann, H. (2009). *Olex2*. Version 1.5-alpha.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. and Jelsch, C. (2012). An improved experimental databank of transferable multipolar atom models – ELMAM2. Construction details and applications. *Acta Crystallographica Section A Foundations of Crystallography* **68**.3, 337–351. DOI: 10.1107/s0108767312008197.

Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, P., Llunell, M., Causà, M. and Noël, Y. (2014). *CRYSTAL14 User's Manual*. University of Torino, Torino, Italija.

Dovesi, R., Orlando, R., Erba, A., Zicovich-Wilson, C. M., Civalleri, B., Casassa, S., Maschio, L., Ferrabone, M., De La Pierre, M., D'Arco, P. and al., et (2014). CRYSTAL14: A program for the *ab initio* investigation of crystalline solids. *Int. J. Quantum Chem.* **114**.19, 1287–1317. DOI: 10.1002/qua.24658.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., J., N., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, D. J. (2009). *Gaussian 09 Revision D.01*. Gaussian Inc. Wallingford CT.

Frontera, A., Gamez, P., Mascal, M., Mooibroek, T. J. and Reedijk, J. (2011). Putting Anion-π Interactions Into Perspective. *Angewandte Chemie International Edition* **50**.41, 9564–9583. DOI: 10.1002/anie.201100208.

Gamez, P., Mooibroek, T. J., Teat, S. J. and Reedijk, J. (2007). Anion Binding Involving π-Acidic Heteroaromatic Rings. *Accounts of Chemical Research* **40**.6, 435–444. DOI: 10.1021/ar7000099.

Gatti, C., Saunders, V. R. and Roetti, C. (1994). Crystal field effects on the topological properties of the electron density in molecular crystals: The case of urea. *The Journal of Chemical Physics* **101**.12, 10686–10696. DOI: 10.1063/1.467882.

Grimme, S., Antony, J., Ehrlich, S. and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**.15, 154104. DOI: 10.1063/1.3382344.

Grounds, O., Zeller, M. and Rosokha, S. V. (2018). Structural preferences in strong anion–π and halogen-bonded complexes: π- and σ-holes vs. frontier orbitals interaction. *New Journal of Chemistry* **42**.13, 10572–10583. DOI: 10.1039/c7nj04843b.

Guillot, B., Enrique, E., Huder, L. and Jelsch, C. (2014). MoProViewer: a tool to study proteins from a charge density science perspective. *Acta Crystallographica Section A Foundations and Advances* **70**.a1, C279–C279. DOI: 10.1107/s2053273314097204.

Heyd, J., Peralta, J. E., Scuseria, G. E. and Martin, R. L. (2005). Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *The Journal of Chemical Physics* **123**.17, 174101. DOI: 10.1063/1.2085170.

Howard, S. T. and Lamarche, O. (2003). Description of covalent bond orders using the charge density topology. *J. Phys. Org. Chem.* **16**.2, 133–141. DOI: 10.1002/poc.584.

Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods usingMoPro. *Journal of Applied Crystallography* **38**.1, 38–54. DOI: 10.1107/s0021889804025518.

Jia, C., Miao, H. and Hay, B. P. (2019). Crystal Structure Evidence for the Directionality of Lone Pair–π interactions: Fact or Fiction? *Crystal Growth & Design* **19**.11, 6806–6821. DOI: 10.1021/acs.cgd.9b01081.

Kepler, S., Zeller, M. and Rosokha, S. V. (2019). Anion–π Complexes of Halides with p-Benzoquinones: Structures, Thermodynamics, and Criteria of Charge Transfer to Electron Transfer Transition. *Journal of the American Chemical Society* **141**.23, 9338–9348. DOI: 10.1021/jacs.9b03277.

Kozuch, S. (2016). Should "anion–π interactions" be called "anion–σ interactions"? A revision of the origin of some hole-bonds and their nomenclature. *Physical Chemistry Chemical Physics* **18**.44, 30366–30369. DOI: 10.1039/c6cp06416g.

Lucas, X., Bauzá, A., Frontera, A. and Quiñonero, D. (2016). A thorough anion–π interaction study in biomolecules: on the importance of cooperativity effects. *Chemical Science* **7**.2, 1038–1050. DOI: 10.1039/c5sc01386k.

Madsen, A. Ø. (2006). SHADEweb server for estimation of hydrogen anisotropic displacement parameters. *Journal of Applied Crystallography* **39**.5, 757–758. DOI: 10.1107/s0021889806026379.

Madsen, A. Ø. and Hoser, A. A. (2015). A simple approach to estimate isotropic displacement parameters for hydrogen atoms. *Acta Crystallographica Section A Foundations and Advances* **71**.2, 169–174. DOI: 10.1107/s2053273314025133.

Meindl, K. (2008). Residual density validation and the structure of Labyrinthopeptin A2, PhD Thesis.

Meindl, K. and Henn, J. (2008). Foundations of residual-density analysis. *Acta Cryst.* **A64**, 404–418.

Molčanov, K., Jelsch, C., Landeros, B., Hernández-Trujillo, J., Wenger, E., Stilinović, V., Kojić-Prodić, B. and Escudero-Adán, E. C. (2018). Partially Covalent Two-Electron/Multicentric Bonding between Semiquinone Radicals. *Crystal Growth & Design* **19**.1, 391–402. DOI: 10.1021/acs.cgd.8b01484.

Molčanov, K. and Kojić-Prodić, B. (2019). Towards understanding π-stacking interactions between non-aromatic rings. *IUCrJ* **6**.2, 156–166. DOI: 10.1107/s2052252519000186.

Molčanov, K., Kojić-Prodić, B., Babić, D. and Stare, J. (2013). Face-to-face stacking of dianionic quinoid rings in crystals of alkali salts of 2,5-dihydroxyquinone in view of π-system polarization. *CrystEngComm* **15**.1, 135–143. DOI: 10.1039/c2ce26119g.

Molčanov, K., Mali, G., Grdadolnik, J., Stare, J., Stilinović, V. and Kojić-Prodić, B. (2018). Iodide···π Interactions of Perhalogenated Quinoid Rings in Co-crystals with Organic Bases. *Crystal Growth & Design* **18**.9, 5182–5193. DOI: 10.1021/acs.cgd.8b00634.

Molčanov, K., Milašinović, V. and Kojić-Prodić, B. (2019). Contribution of Different Crystal Packing Forces in π-Stacking: From Noncovalent to Covalent Multicentric Bonding. *Crystal Growth & Design* **19**.10, 5967–5980. DOI: 10.1021/acs.cgd.9b00540.

Mooibroek, T. J., Gamez, P. and Reedijk, J. (2008). Lone pair–π interactions: a new supramolecular bond? *CrystEngComm* **10**.11, 1501. DOI: 10.1039/b812026a.

Newberry, R. W. and Raines, R. T. (2017). The n→π* Interaction. *Accounts of Chemical Research* **50**.8, 1838–1846. DOI: 10.1021/acs.accounts.7b00121.

Papajak, E., Zheng, J., Xu, X., Leverentz, H. R. and Truhlar, D. G. (2011). Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *Journal of Chemical Theory and Computation* **7**.10, 3027–3034. DOI: 10.1021/ct200106a.

Quiñonero, D., Garau, C., Rotger, C., Frontera, A., Ballester, P., Costa, A. and Deyà, P. M. (2002). Anion–π Interactions: Do They Exist? *Angewandte Chemie International Edition* **41**.18, 3389–3392. DOI: 10.1002/1521-3773(20020916)41:18<3389::aid-anie3389>3.0.co;2-s.

Savastano, M., Bazzicalupi, C., García, C., Gellini, C., López de la Torre, M. D., Mariani, P., Pichierri, F., Bianchi, A. and Melguizo, M. (2017). Iodide and triiodide anion complexes involving anion–π interactions with a tetrazine-based receptor. *Dalton Transactions* **46**.14, 4518–4529. DOI: 10.1039/c7dt00134g.

Schottel, B. L., Chifotides, H. T. and Dunbar, K. R. (2008). Anion-π interactions. *Chem. Soc. Rev.* **37**.1, 68–83. DOI: 10.1039/b614208g.

Shahbazian, S. (2018). Frontispiece: Why Bond Critical Points Are Not "Bond" Critical Points. *Chemistry - A European Journal* **24**.21. DOI: 10.1002/chem.201882161.

Sheldrick, G. M. (2007). A short history ofSHELX. *Acta Crystallographica Section A Foundations of Crystallography* **64**.1, 112–122. DOI: 10.1107/s0108767307043930.

Singh, S. K. and Das, A. (2015). The n → π* interaction: a rapidly emerging non-covalent interaction. *Physical Chemistry Chemical Physics* **17**.15, 9596–9612. DOI: 10.1039/c4cp05536e.

Tang, W., Sanville, E. and Henkelman, G. (2009). A grid-based Bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter* **21**.8, 084204. DOI: 10.1088/0953-8984/21/8/084204.

Tsirelson, V. G., Bartashevich, E. V., Stash, A. I. and Potemkin, V. A. (2007). Determination of covalent bond orders and atomic valence indices using topological features of the experimental electron density. *Acta Crystallographica Section B Structural Science* **63**.1, 142–150. DOI: 10.1107/s0108768106046003.

Vuković, V., Leduc, T., Jelić-Matošević, Z., Didierjean, C., Favier, F., Guillot, B. and Jelsch, C. (2021). *A rush to explore protein-ligand electrostatic interaction energy with Charger.* Submitted for publication.

Vuković, V., Molčanov, K., Jelsch, C., Wenger, E., Krawczuk, A., Jurić, M., Androš Dubraja, L. and Kojić-Prodić, B. (2019). Malleable Electronic Structure of Chloranilic Acid and Its Species Determined

by X-ray Charge Density Studies. *Crystal Growth & Design* **19**.5, 2802–2810. DOI: 10.1021/acs.cgd.9b00033.

Wang, D.-X. and Wang, M.-X. (2013). Anion–π Interactions: Generality, Binding Strength, and Structure. *Journal of the American Chemical Society* **135**.2, 892–897. DOI: 10.1021/ja310834w.

Wick, C. R. and Clark, T. (2018). On bond-critical points in QTAIM and weak interactions. *Journal of Molecular Modeling* **24**.6. DOI: 10.1007/s00894-018-3684-x.

Wilson, J., Maxson, T., Wright, I., Zeller, M. and Rosokha, S. V. (2020). Diversity and uniformity in anion–π complexes of thiocyanate with aromatic, olefinic and quinoidal π-acceptors. *Dalton Transactions* **49**.25, 8734–8743. DOI: 10.1039/d0dt01654c.

Yanai, T., Tew, D. P. and Handy, N. C. (2004). A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **393**.1-3, 51–57. DOI: 10.1016/j.cplett.2004.06.011.

Zhurov, V. V., Zhurova, E. A. and Pinkerton, A. A. (2008). Optimization and evaluation of data quality for charge density studies. *J. Appl. Cryst.* **41**, 340–349.

Cyclopentadienyl‴NiP_3 and Crystal Electrostatic Energy

Abraham, S. and Keve, E. T. (1971). Normal probability plot analysis of error in measured and derived quantities and standard deviations. *Acta Cryst.* **A27**, 157–165.

Bak, B. and Led, J. (1969). Bent bonds in cyclopropane, cyclopropene. Cyclobutane, and cyclobutene. *Journal of Molecular Structure* **3**.4-5, 379–384. DOI: 10.1016/0022-2860(69)87034-1.

Bourhis, L., Dolomanov, O., Gildea, R., Howard, J. and Puschmann, H. (2015a). *olex2.refine*.

— (2015b). *ShelXT*.

Cossairt, B. M., Piro, N. A. and Cummins, C. C. (2010). Early-Transition-Metal-Mediated Activation and Transformation of White Phosphorus. *Chemical Reviews* **110**.7, 4164–4177. DOI: 10.1021/cr9003709.

Dolomanov, O. V., Bourhis, L., Gildea, R. and Puschmann, H. (2009). *Olex2*. Version 1.5-alpha.

Elian, M., Chen, M. M. L., Mingos, D. M. P. and Hoffmann, R. (1976). Comparative bonding study of conical fragments. *Inorganic Chemistry* **15**.5, 1148–1155. DOI: 10.1021/ic50159a034.

Guillot, B., Enrique, E., Huder, L. and Jelsch, C. (2014). MoProViewer: a tool to study proteins from a charge density science perspective. *Acta Crystallographica Section A Foundations and Advances* **70**.a1, C279–C279. DOI: 10.1107/s2053273314097204.

Hoffmann, R. (1982). Building Bridges Between Inorganic and Organic Chemistry (Nobel Lecture). *Angewandte Chemie International Edition in English* **21**.10, 711–724. DOI: 10.1002/anie.198207113.

Hoidn, C. M., Scott, D. J. and Wolf, R. (2020). Transition-Metal-Mediated Functionalization of White Phosphorus. *Chemistry – A European Journal* **27**.6, 1886–1902. DOI: 10.1002/chem.202001854.

Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods usingMoPro. *Journal of Applied Crystallography* **38**.1, 38–54. DOI: 10.1107/s0021889804025518.

Mädl, E., Balázs, G., Peresypkina, E. V. and Scheer, M. (2016). Unexpected Reactivity of [(ν_5-1,2,4-tBu$_3$C$_5$H$_2$)Ni(ν_3-P$_3$)] towards Main Group Nucleophiles and by Reduction. *Angewandte Chemie International Edition* **55**.27, 7702–7707. DOI: 10.1002/anie.201601775.

Madsen, A. Ø. (2006). SHADEweb server for estimation of hydrogen anisotropic displacement parameters. *Journal of Applied Crystallography* **39**.5, 757–758. DOI: 10.1107/s0021889806026379.

Madsen, A. Ø. and Hoser, A. A. (2015). A simple approach to estimate isotropic displacement parameters for hydrogen atoms. *Acta Crystallographica Section A Foundations and Advances* **71**.2, 169–174. DOI: 10.1107/s2053273314025133.

Meindl, K. (2008). Residual density validation and the structure of Labyrinthopeptin A2, PhD Thesis.

Meindl, K. and Henn, J. (2008). Foundations of residual-density analysis. *Acta Cryst.* **A64**, 404–418.

Riesinger, C., Dütsch, L., Balázs, G., Bodensteiner, M. and Scheer, M. (2020). Cationic Functionalisation by Phosphenium Ion Insertion. *Chemistry – A European Journal* **26**.71, 17165–17170. DOI: 10.1002/chem.202003291.

Rigaku Oxford Diffraction (2020). *CrysAlisPro*. Version V1.171.41.93a.

Scheer, M., Balázs, G. and Seitz, A. (2010). P_4 Activation by Main Group Elements and Compounds. *Chemical Reviews* **110**.7, 4236–4256. DOI: 10.1021/cr100010e.

Vuković, V., Molčanov, K., Jelsch, C., Wenger, E., Krawczuk, A., Jurić, M., Androš Dubraja, L. and Kojić-Prodić, B. (2019). Malleable Electronic Structure of Chloranilic Acid and Its Species Determined by X-ray Charge Density Studies. *Crystal Growth & Design* **19**.5, 2802–2810. DOI: 10.1021/acs.cgd.9b00033.

Zhurov, V. V., Zhurova, E. A. and Pinkerton, A. A. (2008). Optimization and evaluation of data quality for charge density studies. *J. Appl. Cryst.* **41**, 340–349.

Conclusions

Nguyen, D., Kisiel, Z. and Volkov, A. (2018). Fast analytical evaluation of intermolecular electrostatic interaction energies using the pseudoatom representation of the electron density. I. The Löwdin α-function method. *Acta Crystallographica Section A Foundations and Advances* **74**.5, 524–536. DOI: 10.1107/s2053273318008690.

A A rush to explore protein-ligand electrostatic interaction energy with Charger

Article published in *Acta Crystallographica*, section **D**: https://scripts.iucr.org/cgi-bin/paper?S2059798321008433.

Vedran Vuković[1], Théo Leduc[1], Zoe Jelić-Matošević[2], Claude Didierjean[1], Frédérique Favier[1], Benoît Guillot[1], Christian Jelsch[1,*]

[1] Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France
[2] Kemijski Odsjek, PMF, Sveučilište u Zagrebu, 10000 Zagreb, Croatia
[*] Corresponding author: christian.jelsch@univ-lorraine.fr

A.1 Abstract

Mutual penetration of electron densities between two interacting molecules complicates the computation of accurate electrostatic interaction energy based on pseudo-atom representation of electron densities. The numerical exact potential and multipole moment (nEP/MM) method is time-consuming since it performs a 3D integration to obtain electrostatic energy at short interaction distances. Nguyen, Kisiel & Volkov [*Acta Cryst.* 2018, A**74**.5, 524–536] recently reported fully analytic computation of the electrostatic interaction energy (aEP/MM). This method performs much faster than nEP/MM (up to two orders of magnitude) and remains highly accurate.

A new program library `Charger` contains an implementation of the aEP/MM method. `Charger` has been incorporated in the `MoProViewer` software. Benchmark tests on a series of small molecules containing only carbon, hydrogen, nitrogen and oxygen atoms show `Charger`'s efficiency in terms of execution time and accuracy. `Charger` is also powerful for a study of electrostatic symbiosis between a protein and a ligand. It determines reliable protein-ligand interaction energies even when both contain sulfur atoms. It estimates easily the individual contribution of every residue to total protein-ligand electrostatic binding energy.

Glutathione transferase (GST) in complex with benzophenone ligands was studied due to availability of both structural and thermodynamic data. The resulting analysis highlights not only which residues stabilize the ligand, but also those that hinder ligand binding from an electrostatic point of view. This offers new perspectives in the search for mutations to improve the interaction between two partners. A proposed mutation would improve ligand binding to GST by removing an electrostatic obstacle, rather than by the traditional increase in the number of favourable contacts.

Synopsis

Embedded within the MoProViewer program, a new code library Charger contains an implementation of the analytic computation of the electrostatic interaction energy based on the multipolar atom. It served to investigate electrostatic interaction energies of benchmark dimers and glutathione transferase-benzophenone complexes.

Keywords

Electrostatics, polarisation, interaction energy, Hansen-Coppens model, protein-ligand interactions

A.2 Introduction

An investigation of intermolecular interactions involves an intelligent electrostatic interaction energy examination. One major obstacle in such energy calculations is how to obtain an accurate representation of total charge distributions of interacting molecules. They are normally associated with theoretical calculation methods such as density functional theory (DFT).

Full DFT calculations on large systems take a prohibitive amount of time. Approximations such as the kernel energy method give comparable results much faster, as they scale better with the number of atoms concerned (Huang *et al.*, 2005; Mandal *et al.*, 2017). This method has been successfully applied on many different biological systems (Massa *et al.*, 2019; and references therein).

Charge density determinations from ultra-high-resolution X-ray crystallography provide an experimental counterpart to theoretical studies. The Hansen & Coppens multipolar pseudo-atom model represents properly an electron density distribution based on an X-ray diffraction experiment (Hansen & Coppens, 1978). A good model necessitates high-quality crystals, which are nearly impossible to produce in some fields. Bio-macromolecular crystallography in particular has this problem. A compromise solution comes from the fact that multipolar parameters of chemically similar atoms are similar. One can reconstruct protein electron density with ease from building blocks (*i.e.* transferable parameters) coming from small molecules.

A parameter transfer from a database to a protein structure takes less than a minute on a modern computer. Transferable multipolar parameter databases draw from experimental (*e.g.* ELMAM2, Domagała *et al.*, 2012) electron densities, augmented with polarizabilities (Leduc, 2019; Leduc *et al.*, 2019), or from theoretical electron densities (*e.g.* INVARIOM, Dittrich *et al.*, 2004; and UBDB, Dominiak *et al.*, 2007).

Missing atom types pose an obstacle to database transfers. Similar atom types can replace them, or (particularly for theoretical databanks) one can obtain its charge density by an additional DFT calculation. Kumar *et al.* (2019) addressed this problem to some extent by extending UBDB with atom types commonly found in drug molecules.

To calculate the electrostatic energy (E_{elec}) remains a challenge even with a fitting electron density model at hand, largely because of the overlap between electron clouds. (Volkov *et al.*, 2004) The challenge of computation coincides with chemist's big interest, since atoms with overlapping electron densities are close enough to form chemically meaningful strong contacts.

One should first divert his attention to a simpler case: atoms that are further away and have negligible density overlap. They can nevertheless contribute significantly to interaction energies, more so when they carry considerable charge. A point charge approximation can describe this kind of interaction well, but the Buckingham approximation (Buckingham, 1959) performs better as it includes multipole moments (MM) along with point charges. These approximations also alleviate time cost concerns associated with large biomolecular systems.

Going back to the more relevant case of atoms that are close together, Volkov *et al.* (2004) developed the exact potential (EP) method for the multipolar electron density model. EP involves a complete numerical 3D integration of the product between a density of one atom and a potential of another atom, and thus handles overlap density directly, albeit slowly. A union of the slow EP method at short distance and the fast MM ap-

proximation at long distance gives rise to the precise and reasonably fast EP/MM method (Volkov *et al.*, 2004). Fournier *et al.* (2009) implemented the same method in the `VMoPro` program of `MoProSuite`.

To improve calculation speed further, Nguyen *et al.* (2018) have proposed an analytical method to evaluate electrostatic interaction energies between charge densities for the Hansen-Coppens model. This method is largely based on previous mathematical developments by Jones and Weatherford (Jones, 1993; and references therein), and Löwdin (1956) and substitutes the numerical EP integrals with the analytical ones. We follow their lead in renaming the original method using numerical integration nEP/MM, and calling the newer one using analytical integration aEP/MM.

Nguyen *et al.* (2018) promise high accuracy with calculation speed improved by two orders of magnitude. D. Nguyen and A. Volkov further analysed aEP/MM precision and proposed a more precise Fourier-based method in a newer paper, (Nguyen & Volkov, 2019) albeit at a certain time cost. They have also recently proposed, in collaboration with P. Macchi, a molecular multipolar moments method easily applicable to crystals, as well as corrections for errors in Ewald summations due to density overlap (Nguyen *et al.*, 2020), following in the footsteps of Bojarowski *et al.* (2016). They tested all methods on small molecules and oligopeptide benchmark systems.

We have adapted the aEP/MM method to create a new code library called `Charger`. It is an independent implementation of the same method with several enhancements geared towards calculation speed. Our molecular visualization and electron density exploration software `MoProViewer` (Guillot *et al.*, 2014) accommodates a simple user interface to `Charger`. It includes a special module for bio-macromolecular structure exploration.

Various benchmark systems containing homo- and heterodimers derived from crystal structures served as `Charger`'s first test. The molecules contained within those systems ranged from water and methanol to Leu-enkephalin and a dodecapeptide, all containing only carbon, hydrogen, nitrogen and oxygen atoms.

Dominiak *et al.* (2009) have already extensively demonstrated that a model from transferred parameters is indeed useful when investigating protein-ligand interactions. Specifically, they used the nEP/MM method to calculate electrostatic interaction energies between a neuraminidase protein and a wide range of its inhibitors.

With this in mind, we wanted to give `Charger` a bigger challenge: glutathione transferase (GST), a 2x240-residue dimeric protein. Schwartz *et al.* (2018) determined four structures of this enzyme, each containing a different benzophenone ligand in its active site. They also measured thermal shifts in the presence of these same ligands, which one can use to determine protein-ligand binding constants and therefore relate to interaction energies (Cimmperman *et al.*, 2008). These ligands make a textbook example to gauge our method: chemically similar, they form hydrogen bonds and similar electrostatic interactions with the host protein. The protein-ligand complexes exhibit distinct thermal shifts for each ligand.

Trying to improve protein-ligand binding is another direction to determine the quality of `Charger` calculations. A typical approach seeks to increase pertinent contacts between them (Thapa & Raghavachari, 2019). The bio-macromolecular module connected to `Charger` can point to residues that hinder ligand binding by poisoning active site's electrostatic environment. We pursued an *in-silico* inquiry in this topic on the GST-benzophenone system.

A.3 METHODS

Charger. Calculations based on quantum chemistry demand the use of computational algorithms. Quantum chemistry often models the world as electron functions that make up orbitals that compose atoms (along with nuclei) that constitute molecules. Computer programs obediently undertake the inhuman drudge of going over the same calculation repeatedly. Our invention, `Charger` partakes in this tradition.

`Charger` is a code library written in C that computes electrostatic interaction energy between molecules. `Charger` works with the Hansen-Coppens (1978) charge density model:

$$\rho_{\text{atom}}(\mathbf{r}) = Z\delta(r) + N_{\text{core}}\rho_{\text{core}}(r) + P_{\text{val}}\kappa^3\rho_{\text{valence}}(\kappa r) + \sum_{l=0}^{l_{\max}} R_l(\kappa' r)\,\kappa'^3 \sum_{m=-l}^{l} P_{lm} d_{lm}(\theta,\phi) \quad (1)$$

This model contains a point-charge nucleus (first term), two spherical and a multipolar term of each atom's electron density (defined in more details in Supplementary material). `Charger` has means to calculate: (1) the electrostatic energy ($E = q_A q_B / r$) of a pair of point charges; (2) the *Exact* electrostatic Potential (*E*P, Φ) from a spherical or multipolar density according to Jones (1993); (3) the energy of a point charge in a potential field ($E = q_A \Phi_B(r_A)$)); and (4) the energy of two interacting electron density (ρ) distributions:

$$E = \iint \frac{\rho_A(\mathbf{r}_A)\,\rho_B(\mathbf{r}_B)}{|\mathbf{r}_A - \mathbf{r}_B|} d\mathbf{r}_A d\mathbf{r}_B \quad (2)$$

(All equations use atomic units.) `Charger` employs the Löwdin's α-function (Löwdin, 1956) to compute this last term when electron density is expressed as a sum of Slater functions, according to the method proposed by Nguyen *et al.* (2018). They dubbed their method *analytical exact potential* (aEP), as opposed to the *numerical* exact potential (nEP) method that Volkov, Koritsanszky and Coppens proposed in 2004 (Volkov *et al.*, 2004).

At short distances, one can calculate through `Charger` the energy as the sum of following terms:

$$E_{\text{Total}} = E_{\text{nn}} + E_{\text{ns}} + E_{\text{nm}} + E_{\text{ss}} + E_{\text{sm}} + E_{\text{mm}} \quad (3)$$

The first term represents nucleus-nucleus energy E_{nn} (point charges Z of the two nuclei, A and B, separated by a distance d):

$$E_{\text{nn}} = \frac{Z_A Z_B}{d}, \quad (4)$$

Two nucleus-electron energy terms E_{ns} and E_{nm} follow (a point charge Z_A in a spherical $\Phi_B(d)$ or multipolar electric potential $\Phi_B(d)\,d_{lm}(\theta_A, \varphi_A)$):

$$E_{\text{ns}} = Z_A \Phi_B(d), \quad (5)$$

$$E_{\text{ns}} = Z_A \Phi_B(d)\,d_{lm}(\theta_A, \varphi_A). \quad (6)$$

The remaining three electron-electron energy terms are E_{ss}, E_{sm} and E_{mm} (two interacting electron density distributions, either spherical or multipolar, defined in Supplementary Material).

A faster calculation at long distances comes from the substitution of nuclear and spherical electron terms in the Hansen-Coppens model with a point charge, in the spirit of the Buckingham approximation (Buckingham, 1959). For a pair of interacting atoms at long distance, we calculate the following sum:

$$E_{\text{Pseudo-Buckingham}} = E_{\text{cc}} + E_{\text{cm}} + E_{\text{mm}} \quad (7)$$

The first term is a simple point-charge energy similar to E_{nn}:

$$E_{\text{cc}} = \frac{q_A q_B}{d} \quad (8)$$

with the charges q of the respective atoms; these charges are defined in the Supplementary material. It substitutes the $E_{\text{nn}} + E_{\text{ns}} + E_{\text{ss}}$ terms from equation (3). The point-multipolar term E_{cm} replaces $E_{\text{nm}} + E_{\text{sm}}$ terms. It is calculated similarly to the E_{nm} term, after substituting Z with q:

$$E_{\text{cm}} = q_A\,\Phi_B(d)\,d_{lm}(\theta_A, \varphi_A) \quad (9)$$

The last term is the same multipolar-multipolar term as in eq. 3 (E_{mm}).

The terms E_{cc} and E_{cm} are similar in nature to the terms E_{nn} and E_{nm}. This allows re-purposing the existing code for a new goal. We call this the *pseudo-Buckingham* approximation, since it does not involve an evaluation of tensor products, as Buckingham proposed. Nevertheless, the E_{mm} term involves a limited number of computations because spherical harmonic functions underlying multipoles are orthonormal.

It is common to refer to the Buckingham approximation as *multipolar moments* (MM) in the literature. We therefore name our combined method analytical Exact Potential / pseudo-Multipolar Moments (aEP/pMM). The program li-

brary Charger provides the functions to calculate energy using both aEP and pMM: library users have the freedom to choose the threshold distances to use the former or the latter.

One important goal of the Charger library is calculation speed. A sizeable calculation speed-up emerges by excluding very long-distance multipolar terms. Charger can use a user-adjustable cut-off scheme based on distances and multipole levels (l).

We established two different cut-off schemes (for "precise" and "rapid" calculations) based on test calculations. These test calculations revolved around typically high multipolar populations: P_{lm} = 0.3 e. The chosen cut-off distances rest on individual multipole-multipole interaction energies for all possible pairs of large multipoles from carbon and hydrogen atoms.

We chose the "precise" cut-offs as those (rounded-up) distances for which the calculated interaction energy is inferior to 5 cal/mol (Table 1). This means that, in an extreme case of very high multipolar populations, the error due to each excluded multipolar term should be lower than 5 cal/mol. The error is generally much lower than that in practice.

The "rapid" scheme cuts deeper: no multipole-multipole interactions calculated beyond 15 Å. The two cut-off schemes share some cut-off criteria concerning high-level multipoles (Table A.1).

We further develop the theoretical background behind these calculations in Supplementary material. We make the aEP/pMM method available in the form of code through the Charger library, and as a ready-made tool to explore electrostatic interaction energy within MoProViewer, our flagship molecular visualisation program (Guillot *et al.*, 2014).

Applying Charger to peptides and proteins . The ultimate proof of a method is in how broad its applications are. One could study many chemical systems with Charger, but our interest lies primarily in the biochemical and bio-structural domain. We started with several small benchmark systems, and proceeded to an exploration of interactions between a protein and several ligands forming complexes of known structure.

We first benchmarked Charger on dimeric systems derived from crystal structures following the lead of Nguyen *et al.* (2018). These systems contain water, glycine, 2-butanone, *N*,*N*-dimethylacetamide, Leu-enkephalin, nonapeptide, decapeptide and dodecapeptide molecules. These benchmarks give some points of comparison between Charger and their code integrated in the XDPROP program. Leu-enkephalin is an opioid neurotransmitter pentapeptide. The nona-, deca- and dodecapeptide consist of leucine and 2-methylalanine residues capped with t-butoxycarbonyl (N-cap) and methyl (O-cap) groups. Leu-enkephalin and these peptides contain carbon, hydrogen, oxygen and nitrogen atom types often found in bio-macromolecules.

The ELMAM2 database does not contain four atom types for 6–8 atoms in the nona-, deca- and dodecapeptide, depending on peptide composition; their electron density was approximated as spherical (and neutral before electroneutrality constraint application). In addition, for comparison purposes, we used the AMBER point-charge model on several benchmark systems for which the database transfer was possible. Charger calculated electrostatic interaction energies based on these AMBER models as well.

We tested Charger further on a larger system: a protein-ligand complex with published experimental structural and thermodynamic data. The glutathione transferase Omega 3S from *Trametes versicolor* (*Tv*GSTO3S) is an intriguing model because crystal structures of this 240-residue enzyme were recently obtained in complex with various hydroxybenzophenone inhibitors (HBPs). Schwartz *et al.* (2018) used thermal shift assays to evaluate thermal stabilization of *Tv*GSTO3S upon HBPs binding.

Studied hydroxybenzophenones have various (in number and position) hydroxyl groups on both phenyl rings (Table A.3). Of the eight HBPs assessed, Schwartz *et al.* had obtained crystal structures of four *Tv*GSTO3S/HBP complexes, which we analysed with Charger (PDB IDs 6f66, 6f67, 6f68 and 6f69).

These protein structures originate from the Protein Data Bank. The PDB files required preparation: an erasure of the C-terminal His-tag (not present in the protein's native form) and

Table A.1. **Cut-offs used in Charger.** `Charger` skips over certain pairs of multipoles at a long distance, making the calculation more robust and faster. We propose two cut-off schemes: one that gives precise results (`MoProViewer` default) and another that gives rapid results. Some criteria are common to both cut-off schemes.

Multipoles	Cut-off for precise calculations / Å	Cut-off for rapid calculations / Å
dipole × dipole	40	
dipole × quadrupole	20	15
dipole × octupole, quadrupole × quadrupole	18	
dipole × hexadecapole, quadrupole × octupole	14	
octupole × hexadecapole	12	
hexadecapole × hexadecapole	10	

The "precise" cut-offs represent (rounded-up) distances for which the calculated interaction energy between a pair of multipoles (stated in the left column, both populations P_{lm} = 0.3) is inferior to 5 cal/mol. The "rapid" scheme cuts all multipole-multipole interactions calculated beyond 15 Å, a distance chosen as reasonable and proven so with subsequent calculations.

an addition of three N-terminal missing residues (methionine-serine-serine) using the Builder tool in `PyMol` (Schrödonger LLC.). We used the `WHAT IF` web server (Hooft *et al.*, 1996) to ascertain hydrogen atom positions that form the most optimal hydrogen bonding network in the protein structure. This procedure is agnostic of any ligands, so we used `MoPro` to add missing hydrogen atoms in ligand molecules (Jelsch *et al.*, 2005).

A manual search for prospective protein-ligand hydrogen bonds followed, involving a visual inspection of protein residues with at least one atom within 5 Å from any atom of the ligand. Hydrogen atom positions of those residues and ligands were adjusted manually where deemed necessary in order to maximize the number and strength of hydrogen bonds. We converted the structure into a `MoPro`-compatible format using `Import2MoPro` (Jelsch *et al.*, 2005).

We developed a multipolar database transfer tool inside `MoProViewer` (Guillot *et al.*, 2014). This tool has an extension for automatic parameter transfer onto protein structures with appropriate charge assignment (+1 to arginine and lysine, −1 to aspartate and glutamate, 0 to others; there were no doubly protonated histidine residues; all atoms of a charged residue share the added charge equally). This new tool transferred experimental multipolar electron density parameters from the ELMAM2 database (Domagała *et al.*, 2012) to the treated protein structures.

Not so long ago, one could not go beyond the described database transfer. In a recent article, we demonstrated that the transferred electron density in complement with theoretical polarizabilities gives access to polarisation energies (Leduc *et al.*, 2019). We built a custom database of theoretical average atomic polarizabilities for ELMAM atoms found in HBPs and proteins. Our computation of electrostatic interaction and polarisation energies of *Tv*GSTO3S-HBP interactions followed the method T. Leduc described in his PhD thesis (Leduc, 2019).

After the transfer of atom polarizabilities from this database, the `MoProViewer` polarisation module served to compute dipoles induced on the ligand by residues near to the ligand and dipoles induced by the ligand on the same residues (until the self-consistency criterion is satisfied). The residues selected for this calculation are at most 3.5 Å away from the ligand (distance between centres of geometry, value selected by visual inspection). They include most of the active site, but not small, negligible contributions of further-away residues that would slow down the polarisation calculation.

With *Tv*GSTO3S/benzophenone charge densities in hand, we could proceed to energy calculations. *Tv*GSTO3S is a homodimeric enzyme,

whose crystal structure contains two independent monomers, each of which binds a benzophenone and a glutathione ligand. Charger calculated total electrostatic interaction and polarisation energies between the two subunits and any available ligands. The calculations cover all possible pairs (first monomer – first ligand, first monomer – second ligand *etc.*). We repeated these calculations with the VMoPro program from MoProSuite to compare the speed and accuracy of our new algorithm based on analytical EP/MM with the established numerical EP/MM method (Jelsch *et al.*, 2005).

Another reason for interest in the *Tv*GSTO3S protein systems are calculations with sulfur atoms. The benchmark systems from Nguyen *et al.* (2018) contain only CHNO atoms. Since their systems do not contain sulfur atoms, the results do not fully confirm the applicability of analytical EP/MM on protein systems. *Tv*GSTO3S contains several cysteine and methionine residues, and the glutathione ligand contains a sulfur atom as well. This allows us to draw more compelling conclusions on the method's applicability.

We furthermore investigated to which degree individual amino acid residues contribute to electrostatic interaction energy of overall protein-ligand binding. Trying to explain these results prompted us to form presumptive hypotheses. We constructed an *in-silico* mutant as a first step towards experimental verification of one such hypothesis: arginine 171 had an unfavourable effect with some ligands.

We chose to mutate it into glutamine in the structure with the PDB code 6f68 (Table A.3). PyMol served to generate the Arg171Glu mutation. The glutamine rotamer was chosen so that its side chain occupies a spatial position similar to that of arginine. We transferred electron density parameters to the resulting mutant structure and realised the same calculations as described above for the wild-type structures.

In all these calculations with Charger, we applied analytical EP for all atom pairs that are closer than 5 Å apart, and pseudo-Buckingham multipolar moments calculation for all remaining atom pairs. We used both cut-off schemes from Table A.1 to determine their quality empirically.

A.4 Results and Discussion

Charger user interface. We integrated the Charger library into MoProViewer and created a user-friendly interface for it. MoProViewer facilitates molecule visualisation and atom selection, while the interface handles parameter reorganisation to a form understandable to Charger.

The options on the Charger interface, such as the selection of atomic constituents (*e.g.* nucleus, core, valence, multipolar...), mirror those already available within the VMoPro program of the MoProSuite. The user can select any set of atomic constituents from the Hansen-Coppens multipolar atom formalism (*i.e.* individual terms from eq. 1), which then comprise the density used in energy calculations. As an example, this enables an inquiry into the interaction of the total protein density with ligand dipoles.

Furthermore, the Charger interface in MoProViewer allows the control over the Buckingham (MM) domain threshold as a fixed distance or a threshold that differs from the sum of van der Waals radii of current atom pairs by a certain distance. The interface also handles the choice of multipolar cut-offs at very long distances, with the "precise" cut-off criteria (Table A.1)set as default for interaction energy calculations.

We further designed a Charger-based calculation module in MoProViewer to provide more insight into protein-ligand interactions. It estimates electrostatic interaction energy contributions of individual residues to protein-ligand interaction energy. It automatically selects residues one at a time and calls Charger to calculate residue-ligand energies. It is conceivable to use this tool to estimate pertinent amino acids at the interface between two proteins. All the parameters that apply to a normal Charger calculation (Buckingham threshold, multipole-multipole cut-offs, atom composition) apply here as well.

Behind the scenes, the program library Charger takes as input molecular structure and electron density information through MoProViewer or some other interface. The electron density comes mostly from atomic tables describing electrically neutral and spherical atoms, along

with some multipolar functions derived empirically to describe non-spherical density.

Once the molecular structure and its electron density model are in place, Charger calculates electrostatic interaction energy between any two groups of atoms, be it a protein and a ligand, one amino acid residue and another, two discrete metal-ligand complexes, or even two ligands attached to the same metal centre. It skips unnecessary calculations – those that result in zero because of fundamental theoretical reason like symmetry or multipoles orthogonality.

Charger performance. The first major point to discuss is of Charger's place in the universe of electrostatic interaction energy calculations. Its method approaches the core problem from a different perspective (analytical rather than numerical integration). It should therefore give similar results as other (numerical) methods that try to compute the same quantities. The challenges posed are many, from small molecules to large protein-ligand complexes. Charger promises faster calculation in all of these cases – a claim we scrutinize after dealing with calculation accuracy.

We tested Charger's analytical exact potential/pseudo-multipolar moments (aEP/pMM) method accuracy against the established numerical nEP/MM method available in VMoPro. In the case of benchmark molecules, we also compared it to available literature data (Nguyen *et al.*, 2018).

The energy results for benchmark calculations are very similar for the standard nEP/MM and our aEP/pMM method (Figure A.1, Tables A.2 and S1). The differences are within 0.2 kcal/mol for the benchmark dimers, and are presumably due to numerical integration errors in the nEP/MM method. The energy values are also largely comparable to those available from Nguyen, Kisiel and Volkov (the aEP/MM method), despite the fact that our calculations relied on an electron density models with a different set of multipolar parameters. The biggest relative differences come from dimers whose interaction energy value is ±10 kcal/mol, which one would expect from significantly different underlying models. Furthermore, Charger calculations with AMBER03 charges (Duan *et al.*, 2003) used as spherical valence charges in the context of a Hansen-Coppens model agree well for almost all tested dimers (Tables 3 and S2). These results are comparable to those by Kumar *et al.* (2014), who performed an in-depth comparison between AMBER, Hansen-Coppens and several other models on the S66 and JSCH-2005 datasets. The agreement is however not so good when using AMBER charges as point charges, which yield much lower energy values due to the missing penetration energy.

Calculation times for aEP/MM and aEP/pMM methods are also of the same order of magnitude (Table S3). Charger appears faster at first glance, particularly for small systems with many close atoms. A test under exactly the same conditions (same processor and molecular model) is needed before we can draw more convincing conclusion. The discussion on calculation speed comparison between aEP/pMM and nEP/MM is further down.

We wanted to establish whether the results of the aEP/pMM method in Charger and the nEP/MM one from VMoPro are comparable. We used the same benchmark dimer energies (Table S1) along with those between glutathione transferase (GST) and its glutathione and hydroxybenzophenone ligands (Tables S4 with electrostatic and S5 with electrostatic+polarisation energies).

Figure A.2 presents the Bland-Altman plot of all these energies, using nEP/MM energies from VMoPro as reference. The y-axis presents the difference between aEP/pMM and nEP/MM results, while the x-axis contains the average of the two. The graph thus presents the difference in results between the two methods as a function of the magnitude of the energy results. It gives a feeling for how different the two methods are for large or small interaction energy (the result of both methods).

Results are on average below the reference VMoPro-nEP/MM method (by 0.02 kcal/mol for 'precise' and 0.06 kcal/mol for 'rapid' calculations), and most fall within two standard deviations (±0.39 kcal/mol for 'precise' and ±0.60 kcal/mol for 'rapid' calculations). Rapid calculations give values that are slightly more dissimilar, but are sometimes much closer to the mark.

Concerning the accuracy of the values, there is no direct comparison between aEP/pMM and external methods here. Leduc *et al.* (2018) compared the nEP/MM method with Symme-

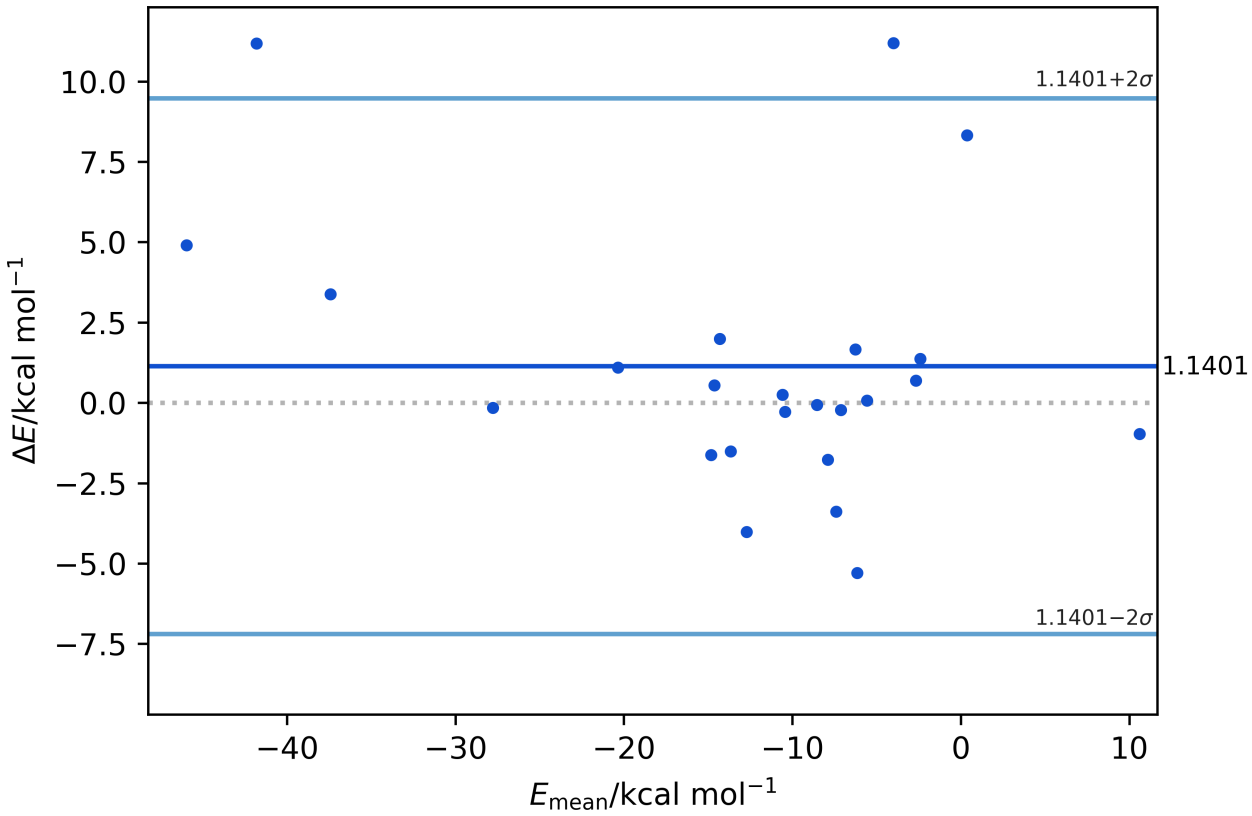


Figure A.1. Bland-Altman plot to assess the difference between `Charger` electrostatic interaction energy calculations and those available from Nguyen *et al.* (2018) on benchmark systems. It depicts the difference in results between *Charger* aEP/pMM (precise cut-off, reported in Table A.1) and the literature reference (`XDPROP` aEP/MM, used as reference: $\Delta E = 0$, dotted grey line) as a function of the magnitude of results. E_{mean} refers to the average value between the `XDPROP` and `Charger` results. Sigmas refer to standard deviations of ΔE values. The values from two Leu-enkephalin and two nonapeptide dimers were omitted for reasons explained in footnotes † and †† of Table S1.

try Adapted Perturbation Theory (SAPT) energies obtained from high-level theoretical calculations on the S66x8 set (containing 66 small molecule dimers at 8 different distances). The correlation was strong (R^2 = 97.0%, slope 1.05; intercept set to zero) without adding further empirical parameters. The RMSD for the whole dataset was estimated to 1.4 kcal/mol. Taken together with the comparison between aEP/pMM and nEP/MM, one should report the energy values in kcal/mol up to the first decimal point, while taking care not to over-interpret small differences in estimated energies.

On the other hand, the availability of the two independent protein monomers gives some insights on the variability of the computed energy due to the 3D structure itself. The relative difference $\sum|\Delta E| / \sum|\langle E\rangle|$ of the ligand/glutathione energy values in Table S5 is typically 20%.

The variability of electrostatic properties obtained by the use of different pseudoatom multipolar databases was verified by Bak *et al.* (2011). The E_{elec} electrostatic energies of crystallographic dimers for a series of amino acids and dipeptides were computed using charge densities transferred from ELMAM2, INVARIOM and UBDB electron density databases and were complemented by first principle calculations (MP2 or DZP). The correlation coefficient of the E_{elec} values from the different calculations was in the 0.79 to 0.99 range and the observed relative difference

Table A.2. **Electrostatic interaction energies of glycine···glycine dimers (in kcal mol^{-1}) for interactions in the benchmark glycine system.**

Program	Calculation	Gly1	Gly2	Gly3	Gly4	Gly5	Gly6
Charger	precise *	−27.9	−7.2	−19.8	−35.7	10.1	−5.5
Charger	rapid *	−27.9	−7.2	−19.8	−35.7	10.1	−5.5
VMoPro	nEP/MM	−27.8	−7.3	−19.8	−35.9	10.1	−5.6
XDPROP	NKV paper**	−27.7	−7.0	−20.9	−39.1	11.1	−5.6
Charger	Amber spherical charges ($q = P_{\mathrm{val}} - N_{\mathrm{val}}$ **	−23.2	−8.9	−16.7	−30.7	3.8	−4.5
Charger	Amber Point charges	−12.0	−0.4	−11.7	−18.5	6.7	−2.3
	δ (NKV) / %	0.7	2.9	−5.3	−8.7	−9.0	−1.8
	δ (Amber sph.) / %	20	−19	19	16	166	22
	δ (Amber point ch.) / %	133	1700	69	93	51	139

δ refers to relative discrepancy with the *Charger* precise values as reference.

$\sum|\Delta E| / \sum|\langle E\rangle|$ was about 24% for energy values issued from the three databases.

The GST structure contains one cysteine and several methionine residues, and the glutathione ligand is a tripeptide (L-glutamyl-L-cysteinylglycine). All of these moieties contain sulfur atoms. The aEP/MM method has not been tested on sulfur atoms to the best of our knowledge (Nguyen *et al.*, 2018; Nguyen & Volkov, 2019). Figure 2 shows absolute discrepancies between VMoPro-nEP/MM and Charger-aEP/pMM results, all of which are lower than 1 kcal/mol. The largest differences come from calculations whose results were the largest (over 150 kcal/mol), corresponding to relative differences of about 1%. These observed differences are attributable to the differences of methods used to computes these energies. In addition, these results show that Charger (and aEP in general) thus appears suitable for calculation with CHNOS atoms found in protein structures.

Tables S3 and S6 contain wall-clock times required to calculate all the interaction energies described thus far on an Intel® Core™ i7-8700 (3.20 GHz) processor. We estimate heavy computation load as the amount of close atom pairs in the calculation, since atoms that are close together elicit the more time-consuming aEP or nEP calculation. Calculation speed-up in Charger portrayed as a function of close atom pairs (less than 5 Å apart) shows up on Figure A.3 for GST-ligand dimers and Figure S1 for all examined dimers.

We observe a significant (30–80 times on average) increase in speed of Charger analytical calculations compared to VMoPro numerical integration for GST-ligand dimers. In the case of benchmark systems, the improvement is even larger (up to 150–200 times). This improvement increases with the number of close atom pairs in the calculation. Closer atoms require more dedication to calculate all interaction integrals, and this afflicts VMoPro-nEP/MM computational time much more severely.

This does not hold true when there are no close atom pairs. Calculating the interaction energy between a GST monomer and the hydroxybenzophenone ligand located in the binding site of the other monomer takes longer for Charger (Table S6). In this case, VMoPro is faster (up to 10 times, or 1-3 seconds in real terms) probably because it includes a proper implementation of the Buckingham potential and Charger includes a pseudo-Buckingham approach.

One could therefore imagine that VMoPro would be faster even when the ligand has few close interactions (i.e. few close pairs of atoms) and many far interactions. The GST monomer and the glutathione ligand from the other monomer's binding site provide such a case: there are as few as 6 close atom pairs per 10 000. Here, Charger is already 4.5x to 15x faster (precise/rapid calculation) than VMoPro-nEP/MM because the analytical calculation for close atom pairs is preponderant and much faster than numerical calculation.

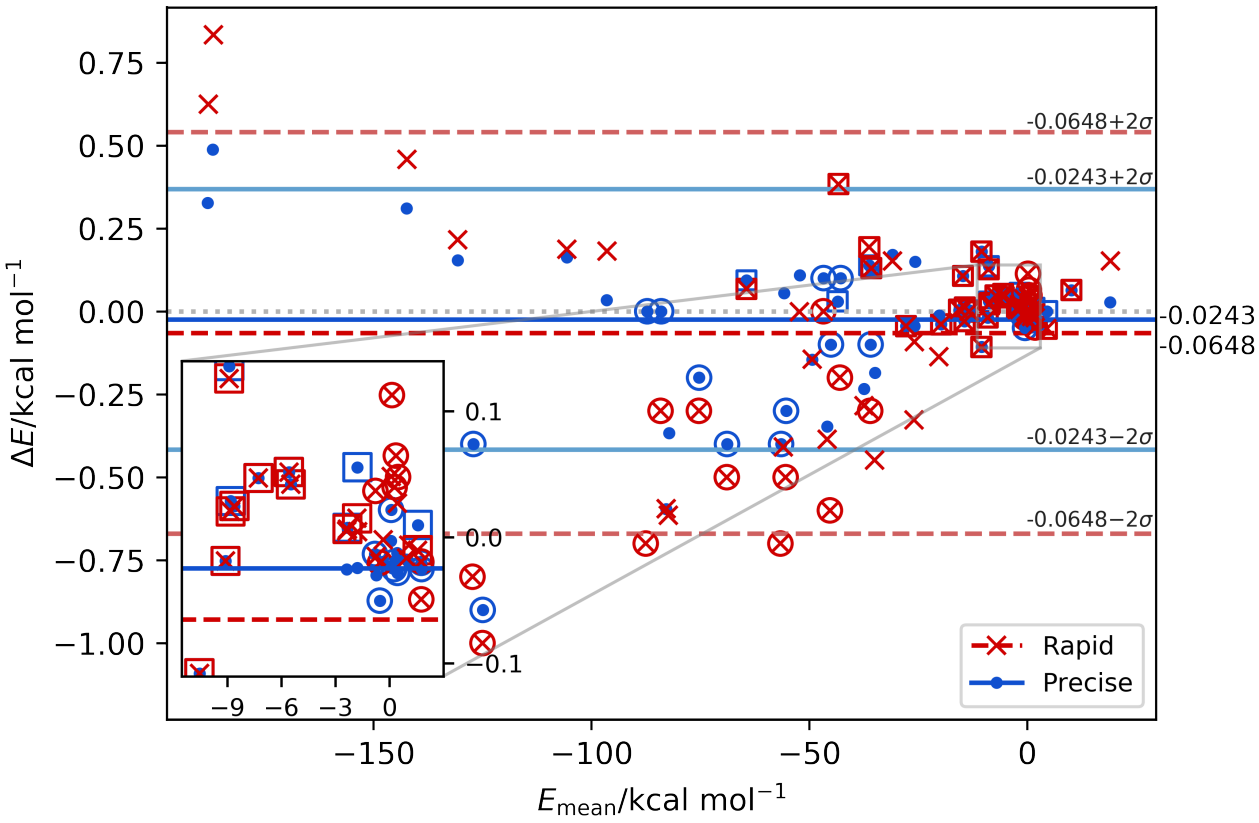


Figure A.2. **Bland-Altman plot to assess the accuracy of `Charger` electrostatic interaction energy calculations.** Bland-Altman plot to assess the accuracy of *Charger* electrostatic interaction energy calculations. It depicts the difference in results between new methods (`Charger` aEP/pMM, rapid/precise) and an old one (`VMoPro` nEP/MM, used as reference: $\Delta E = 0$, dotted grey line) as a function of the magnitude of results. E_{mean} refers to the average value between the `VMoPro` and `Charger` results. Sigmas refer to standard deviations of ΔE values for the respective calculations. The graph contains energies from benchmark dimers (in squares) and GST-HBP interaction calculations (in circles: polarised density, others: non-polarised density). The latter interaction energies include electrostatic and electrostatic + polarisation energies. The inset magnifies the area around the origin for better assessment of points clustered in that region. "Precise" and "rapid" refer to cut-off criteria, reported in Table A.1.

When comparing cut-offs, "rapid" calculations with a more severe threshold (15 Å for low-order multipoles) are about 20% faster for small systems with many atoms in close contact, such as Leu-enkephalin dimers (77 x 77 atoms, Figure S2). Larger systems with few or zero atoms in close contact run 1.5 to 3 times faster (GST monomer with benzophenones: 3706 x 26 or 27 atoms; or with glutathione: 3706 x 36 atoms; or a dimer with twice as many atoms).

"Rapid" calculation does not suffer from a large loss of accuracy (Figure A.2). On our processor, the saved time in real terms amounts to a tenth of a second for small systems and between two and eleven seconds for larger systems. The benefits of using a 15 Å threshold for single point calculations on small to mid-size proteins are modest, but a large-scale calculation (larger protein, database search) would profit significantly.

`MoProViewer` allows users to select two groups of interacting atoms for the energy computation. To explore energy contributions of individual amino acid residues to overall protein monomer-ligand interaction energy, `MoProViewer` silently selects each residue and then uses `Charger` to compute its interaction energy with the chosen ligand. This feature has a very

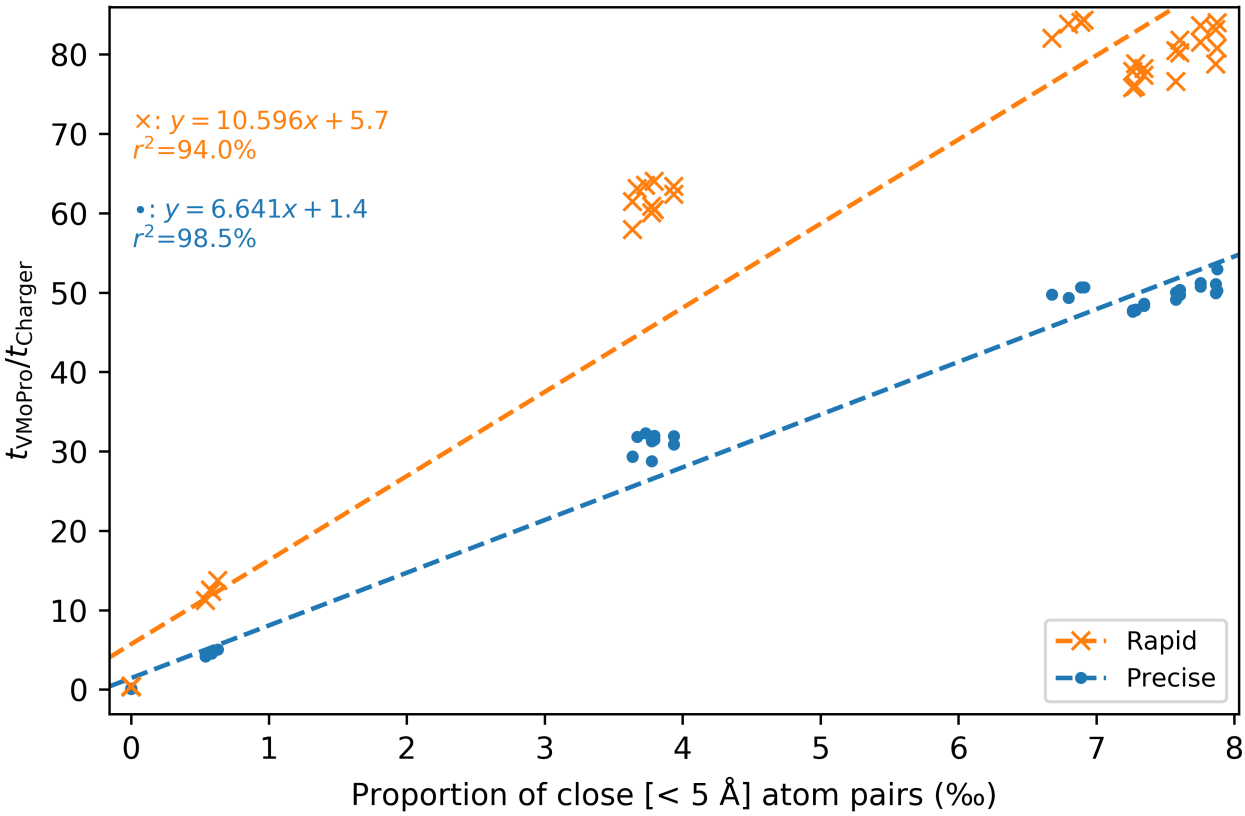


Figure A.3. **Computation speed enhancement.** Improvement in *Charger* calculation speed as compared to VMoPro-EP/MM. The graph focuses only on data from protein/ligand systems, which have few close atom pairs. The x-axis represents the proportion of close atom pairs (in promille, ‰), *i.e.* the number of atoms closer than 5 Å apart divided by the number of all atom pairs. Precise and rapid cut-off criteria are noted in Table A.1. The figure S1 in the Supplementary material shows the full graph with protein/ligand and benchmark calculations.

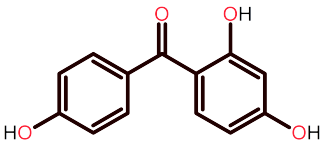
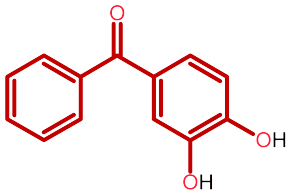
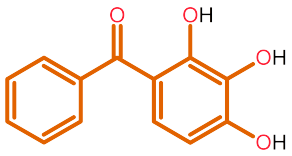
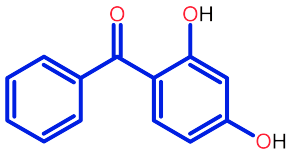
small overhead of 1% in computation time compared to calculating the total energy between the ligand and a whole monomer.

It is possible to verify these individual residue results with VMoPro's nEP/MM method. It already has capacity to calculate several interaction energies arising from multiple atom selections within one structure. Accessing this capability from MoProViewer requires implementing a function to generate the appropriate VMoPro scripts from MoProViewer selections, which was not undertaken. We expect such a calculation to be nearly as fast as a calculation between the whole protein and the ligand using VMoPro.

Interaction analysis with Charger. Having established Charger as a reliable tool to calculate electrostatic interaction energy quickly even for bio-macromolecules, we turn our attention to investigating these same molecules. The possibility to correlate protein-ligand energies with experimental results (thermal shift assays) peaked our interest.

In the study by Schwartz *et al.*, the crystallographic results seem to have good agreement with the thermal shift assay results in solution (*i.e.* observed differences in denaturation temperatures, ΔT_d). It was not possible to obtain the crystallographic structures of complexes with the hydroxybenzophenones that had the lowest ΔT_d values. On the contrary, the HBPs, with the highest ΔT_d, seemed to be the well bound in the active site. Investigating individual residue-ligand concept became so appealing that we created a special module in MoProViewer dedicated to it.

Table A.3. **Thermodynamic results obtained for *Tv*GSTO3S with hydroxybenzophenone inhibitors by thermal-shift assays (Schwartz *et al.*, 2018).** TThe PDB IDs of the crystal structures of *Tv*GSTO3S/HBP complexes are indicated along with experimental thermal shifts (ΔT_d). The colouring of ligands' carbon skeletons reflects the colouring on graphs and structure models throughout this paper.

Compound	Structure	ΔT_d	PDB ID
2,4,4'-trihydroxybenzophenone (2,4,4'-HBP)		2.9	6f68
3,4-dihydroxybenzophenone (3,4-HBP)		4.4	6f67
2,3,4-trihydroxybenzophenone (2,3,4-HBP)		5.0	6f69
2,4-dihydroxybenzophenone (2,4-HBP)		5.7	6f66

First, a short introduction to the protein-ligand complex is in order. `Charger` determined the electrostatic energy of the interaction between the glutathione transferase Omega 3S from *Trametes versicolor* (*Tv*GSTO3S) and four hydroxybenzophenone inhibitors (HBPs, Table A.3). *Tv*GSTO3S is a dimeric enzyme; each monomer contains an active site with a glutathione binding site (G-site) and a hydrophobic pocket for the electrophilic substrate (H-site) (Schwartz *et al.*, 2018).

Hydroxybenzophenones sit in the H-site where 2,4,4'-HBP has a distinct conformation from that of 2,4-, 2,3,4- and 3,4-HBPs (Figure A.4). Thermal shift assays show that these latter three had greater thermal stability compared to the 2,4,4'-HBP complex. The crystal structure of this complex (PDB ID 6f68) shows higher thermal motion and the disordered character of this ligand (a main conformation was modelled at occupancy factor 0.79) is consistent with its low thermal shift of only ΔT_d = 2.9 K.

Schwartz *et al.* (2018) did not establish a quantitative correlation between the thermal shifts and the ligand binding strengths. Interestingly, these thermal shifts and `Charger` energies for monomers and their own ligands are well correlated (correlation coefficient values generally above 85%, Figures A.5 and S3 and notes under Table S7). The correlations of thermal shifts with interaction energies of one monomer and the ligand bound in the second monomer are much poorer (correlation coefficient values generally below 40%). Stronger correlations in the first case are expected, since the ligand within the binding site normally has a stronger influence on overall protein-ligand complex stability.

The order of thermal shift values measured for the considered complexes with benzophenones matches nearly perfect that of electrostatic ligand

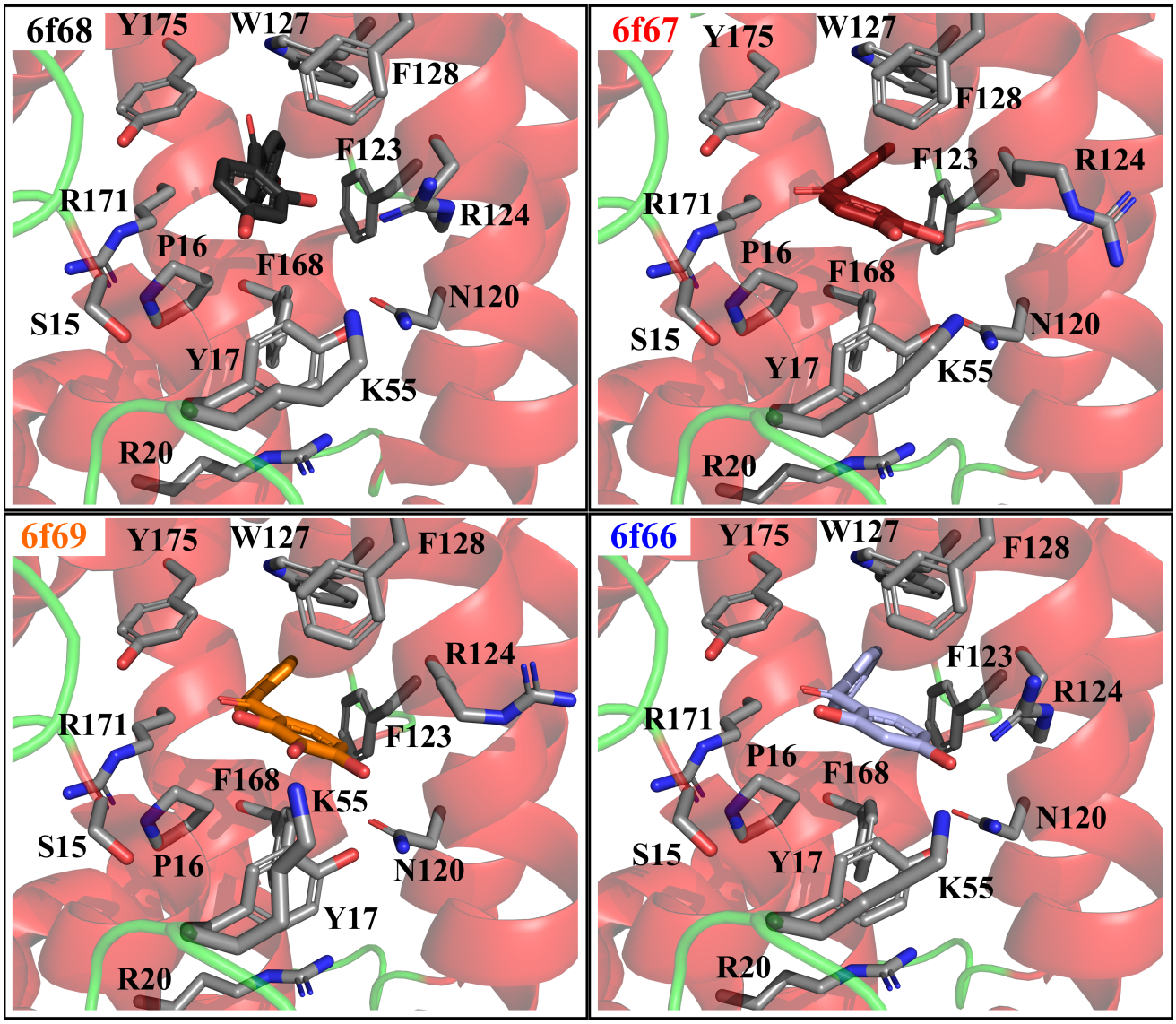


Figure A.4. **Binding of benzophenones (2,4; 3,4; 2,4,4′ and 2,3,4-HBPs) in *Tv*GSTO3S H-site (Schwartz *et al.* 2018, PDB IDs 6f66, 6f67, 6f68 and 6f69 respectively).** *Tv*GSTO3S H-site is a well-delineated deep cavity. Polar residues (Tyr17 and Arg124) at the cavity entrance close to the glutathione binding site and the catalytic residue Ser15 complete the H-site. Side chains and benzophenone ligands are represented as sticks. The conformational difference between the ligand in the 6f68 structure and other ligands resides in the dihedral angles between the carbonyl plane and the aromatic rings (Schwartz *et al.* 2018). The ligand colouring follows the convention from the Table A.3.

binding energies computed with `Charger`. Trend line slopes (around −6.5 kcal/mol/K for monomers and −12 kcal/mol/K for the dimer) indicate that we should detect a different thermal shift by at least 1 K if we calculate a difference of around −6 kcal/mol/monomer. As expected, these numbers are higher when considering polarisation (around −9 kcal/mol/K for monomers and −17 kcal/mol/K for the dimer). An experimentally observable difference in the thermal shift of 1 K should arise for a small structure modification (like a slight chemical modification of a ligand or a residue mutation) that changes the interaction energy per monomer by roughly −6 or −9 kcal/mol (without or with polarisation). This could serve as a rough guide for a ligand design (*e.g.* with docking) or a protein

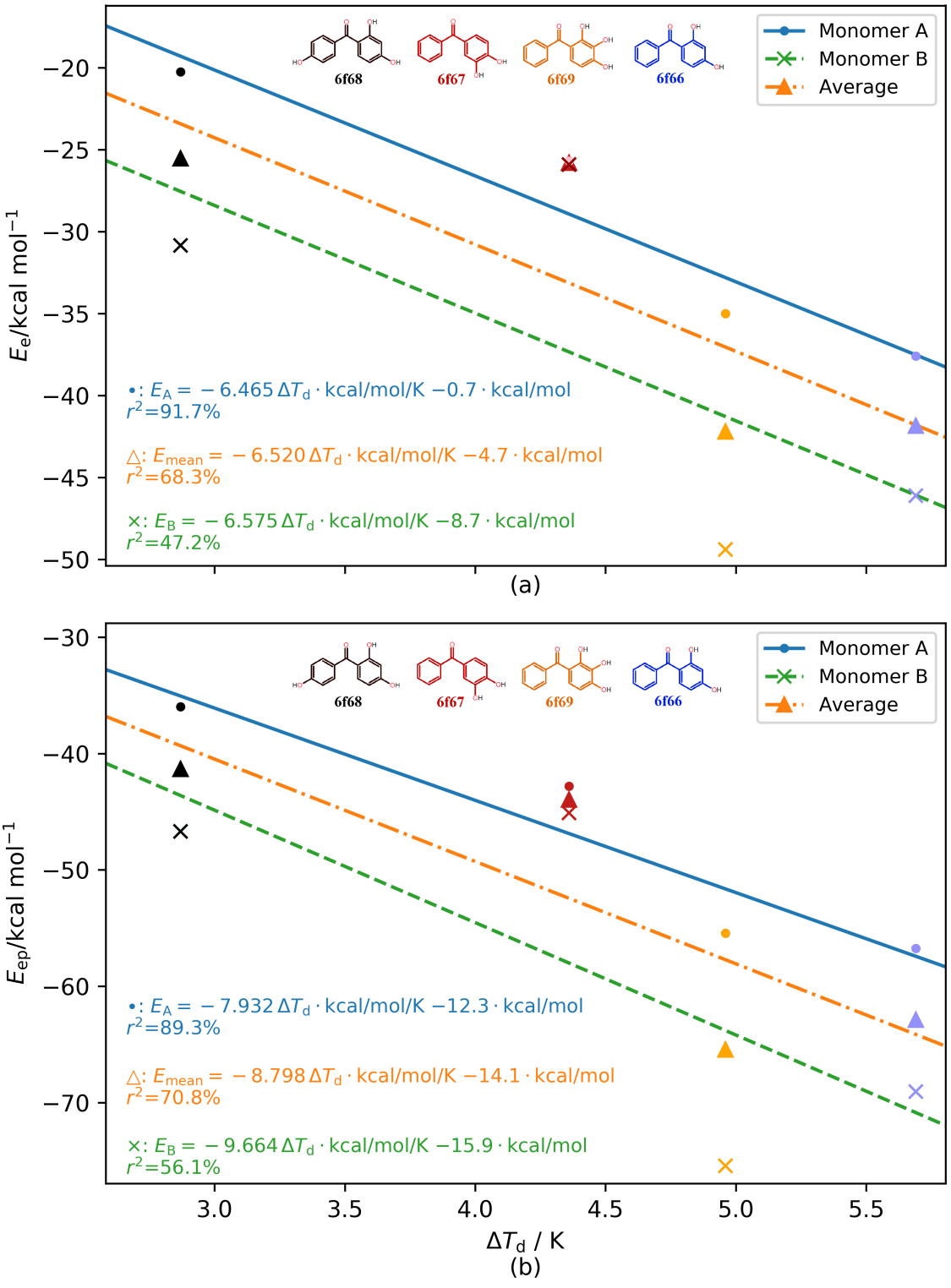


Figure A.5. **Relationship between estimated experimental thermal shifts and computed energies in GST monomers.** Relationship between computed electrostatic interaction energies (a: without polarisation, b: with polarisation) from `Charger` calculations and estimated experimental thermal shifts for glutathione transferase and four ligands. The quantities E_e and E_{ep} represent interaction energies between each monomer and the ligand bound in its active site, along with their mean values. Thermal shift values given in Table S7.

mutation study. Thermal shift analysis is a commonly used tool used in lead compounds identification (screening) during drug discovery and to a lesser extend in lead optimization (Pantoliano *et al.*, 2001; Xing *et al.*, 2019).

A typical analysis of a macromolecular complex's three-dimensional structure highlights assumed interactions that favour macromolecule – small molecule adhesion. This analysis involves visual inspection and distances measurements. Schwartz *et al.* (2018) performed this "manual analysis" in the complexes of *Tv*GSTO3S with benzophenone ligands. They assumed that H-site of *Tv*GSTO3S is suited to accommodate HBPs, due to the presence of polar residues (Tyr17 and Arg124), aromatic side chains (Phe123, Trp127, Phe128, Phe168 and Tyr175) and the aliphatic part of Arg171 (Figure A.5).

`Charger` calculates the contribution (favourable or unfavourable) of each individual residue to the total interaction energy between a ligand and *Tv*GSTO3S. Observing interaction energy contributions of all 240 individual residues reveals that the two monomers are roughly similar, with a remarkably high degree of intra-monomer correlation (Table S8). The analysis of the structures revealed indeed no difference between the subunit main chains (mean RMSD of 0.12 Å). There is even a pronounced correlation between per-residue interaction energies for monomers hosting different ligands, except for the ligand with a 4'-hydroxyl group (6f68, 2,4,4'-HBP, Table S8). The uncommon position of the 4'-hydroxyl ligands within the binding site can explain these weaker correlations.

We went through the individual binding contributions to find the most interesting residues – those with the most pronounced electrostatic interaction energy. We realised that many residues have a minor contribution (Figure S5). Taken at the protein scale, these minor favourable and unfavourable contributions roughly balance themselves out.

We then used a simple filtering criterion: at least one ligand / residue electrostatic interaction energy surpasses ±3 kcal/mol, corresponding roughly to a 3σ standard deviation (Figure A.6). As expected, the retained residues included most of the residues found by visual inspection (Tyr17, Arg124, Trp127, Phe128, Phe168, Arg171 and Tyr175), with notable exception of Phe123. Interestingly, `Charger` also identified residues not suspected from a visual search: Pro16, Arg20, Lys55 and Asn120. We inspected these residues closely, and detail here the interactions of Pro16, Tyr17, Phe168 and Arg171. Other residues have similar energy profiles (Figure A.6) and explanations on the atomic level.

Tyr17 contributes favourably to 2,4-, 3,4- and 2,3,4- HBPs binding, and not significantly to 2,4,4'-HBP. In the first three cases, Tyr17 side chain binds the benzophenone 4-hydroxyl group *via* a hydrogen bond. The particular 2,4,4'-HBP conformation disallows the formation of an equivalent hydrogen bond (Schwartz *et al.*, 2018), so the contribution of this residue becomes negligible.

The contribution of Phe168 is favourable in all cases due to C-H···π interactions. Contrary to the previous case, it contributes much more to 2,4,4'-HBP binding due to an additional hydrogen bond between the Phe168 carbonyl group and the 4'-hydroxyl group (see Figure S6 and section Electric field lines topography analysis on Arg171Gln mutant in SI). The 2,4,4'-HBP ligand has a particular binding mode, different from the other three benzophenones that do not have a 4'-hydroxyl group. The binding of this ligand is not favoured by several nearby residues (Arg20, Asn120 and Arg171), as their interaction energies are unfavourable (sum of electrostatic interaction energies for the cited residues is +10.1 kcal/mol, averaged between two monomers, Figure A.6).

However, the 2,4,4'-HBP forms a 4'-O-H···O=C hydrogen bond with the Phe168 carbonyl group. The O···O distances of 2.75 and 2.66 Å (monomer A and B) are shorter than 3 Å, which enables to classify them as strong H-bond (MacLeod & Rosei, 2014). The balance between these unfavourable contributions and this strong H-bond specific to the 4'-OH substitution might induce this ligand to adopt the specific binding mode observed in the crystal structure.

The calculated electrostatic interaction energy reflects this trend (−16.2 kcal/mol for the 2,4,4'-HBP, average over two binding sites, compared to −4.8±0.9 kcal/mol for the 4'-hydrogen ligands, average and standard deviation for three ligands in

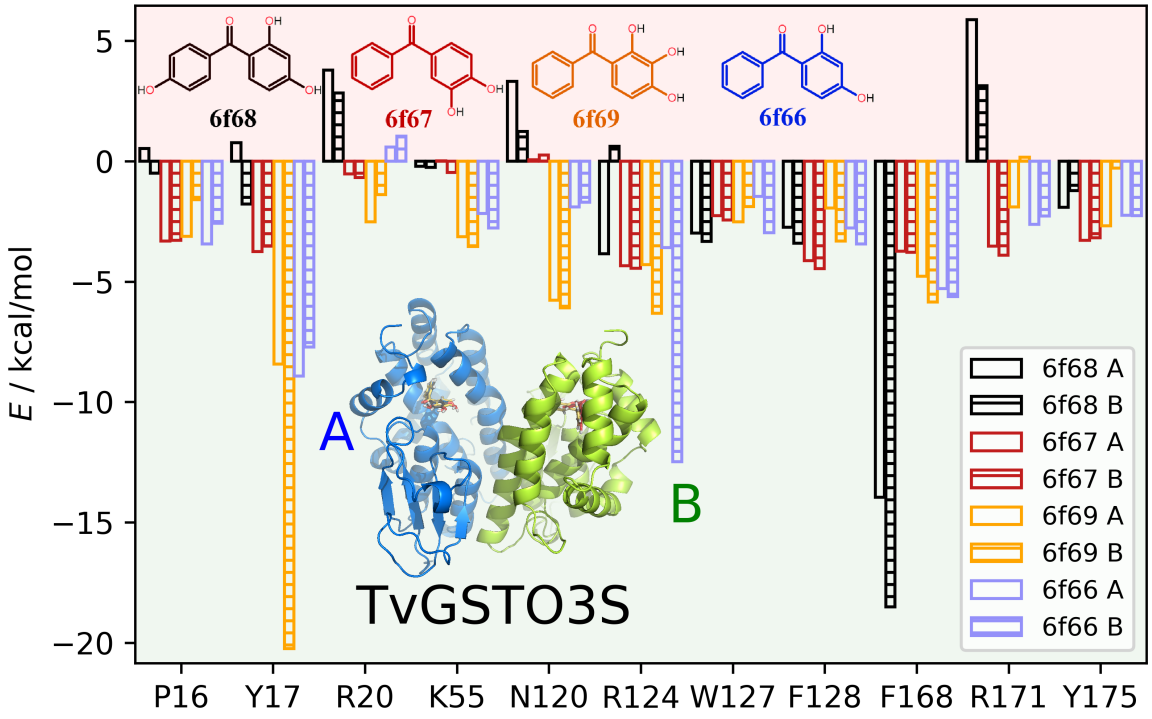


Figure A.6. **Energy contribution of protein residues to ligand binding.** Negative electrostatic (without polarisation) interaction energy contributions of individual residues to GST-ligand binding for the two GST monomers (A and B). Residues with no energy contribution larger than ±3 kcal/mol are omitted for clarity. The use of negative energy means that positive numbers on the *y*-axis indicate favourable binding contribution (green background), and negative numbers indicate unfavourable binding (red background). Legend refers to PDB codes given in Table A.3. The ligands follow the ascending order of their thermal shifts.

two binding sites). The −4.8±0.9 kcal/mol roughly correspond to the interaction energy between a ligand aromatic ring and the Phe168 phenyl ring, an interaction present for all four ligands. The described strong hydrogen bond in the case of 2,4,4'-HBP could be responsible for the additional energy of roughly −11 kcal/mol. One should keep in mind the desolvation penalty – the effect of hydrogen bonds on ligand binding strength is often strongly attenuated due to almost iso-energetic hydrogen bonds with the bulk water solvent.

`Charger` pointed towards an amino acid not deemed significant by visual inspection in the study by *Schwartz et al.*: Pro16. This residue interacts favourably with 2,4-, 3,4- and 2,3,4-HBPs (−5.3±0.9 kcal/mol, mean±s.d.) and interacts more weakly with 2,4,4'-HBP (−0.8 kcal/mol in one binding site, and −2.4 kcal/mol in the other). A closer contact between Pro16 and 2,4-, 3,4- and 2,3,4-HBPs might in part explain these differences. The shortest interatomic distance (from proline C_δ to a ligand oxygen atom) is 4.5 Å for 2,4,4'-HBP and 3.7 Å for the three other benzophenones.

The visual inspection of the *Tv*GSTO3S/HBP crystal structures predicts weak but favourable C-H···π and van der Waals interactions between the HBP ligands and the aliphatic part of the Arg171 side chain (Schwartz *et al.* 2018). A different interpretation appears under "`Charger`'s microscope". When considering only permanent multipoles, indeed, it predicts a strong favourable inter-

action energy between Arg171 and 2,4-, 3,4- and 2,3,4-HBPs (−2.6±1.1 kcal/mol, mean±s.d). This interaction appears similar to the one with Pro16, but slightly stronger. Surprisingly, and unlike in the Pro16 case, it detects an unfavourable interaction energy between Arg171 and 2,4,4'-HBP of about +4.5 kcal/mol (Figure A.6).

Calculation and analysis of molecular dipole moments, using `VMoPro` and `MoProViewer`, provide some explanations on how Arg171 affects ligand binding. Dipole moments of 2,4,4'-HBP and Arg171 are at angles of 63 and 87° in the two monomers (Table S9, Figure A.7.a). The positive charge distributions of Arg171 and 2,4,4'-HBP seem to interfere. This renders the contact between the ligand and the amino acid side chain unfavourable. On the other hand, the antiparallelism between the dipole moment of 2,4-, 3,4- and 2,3,4- HBPs and that of Arg171 (angles of around 150 to 170°, Table S9, Figure A.7.b) could signal a favourable Keesom interaction (the interaction between permanent dipoles, Keesom, 1921).

Furthermore, there is a narrow solvent-accessible cavity near the surface of Arg171 side chain extremity (Fig. S7). This cavity is likely to contain a counter-ion that would alleviate the influence of Arg171. The counter-ion would be located further away from the ligand than Arg171, so its influence should be smaller than that of the Arg171 charge.

The addition of polarisation density supports the general conclusion. It largely compensates for the negative effect of the Arg171 on 2,4,4'-HBP binding (electrostatic interaction energy of about +0.8 kcal/mol), and improves the standing of other ligands (interaction energy of −7.9±1.6 kcal/mol). Though Arg171 does not hinder 2,4,4'-HBP binding severely, as one would conclude without polarisation density, a mutation of this residue could still improve ligand binding (see next paragraph).

Interaction prediction with `Charger`. Results of the detailed investigation into GST-benzophenone binding prompted us to consider research perspectives. One promising offshoot is a mutation study on GST.

We have tried to improve the 2,4,4'-HBP binding by an *in-silico* mutation. Arginine 171 intrigued us through its seemingly unfavourable dipole interaction with this ligand. An attenuation of the arginine charge and polar contributions could establish whether that contribution is responsible for weakly unfavourable binding energy.

We first considered alanine, but dismissed it since replacing the large arginine residue by a small side chain one leaves a large void in the active site. We looked for a suitable neutral amino acid of approaching size. We supposed that mutating Arg171 to a glutamine should have an effect on dipole moment due to the introduction of a relatively electronegative oxygen atom and to the loss of the formal charge of arginine.

The dipole moment of an *in silico* mutated Gln171 is much smaller than that of Arg171. The dipole orientation relative to 2,4,4'-HBP remains similar to the one found in wild-type Arg171 (Figure A.7.c,d, Table S9). Furthermore, the addition of polarisation density shows how the mutation affects the entire binding pocket. The negative effect of the wild-type Arg171 on 2,4,4'-HBP binding diminishes when taking polarisation into account (average interaction energy of about +0.8 kcal/mol instead of +4.5 kcal/mol without polarisation). The Arg171Gln mutant remains favourable: the Gln171·2,4,4'-HBP interaction energy (with polarization) improves to around −2.1 kcal/mol (Table S9). This improvement likely comes from neutralising the dipole effect, largely due to the glutamine lack of charge. An analysis of electric field lines underscores this interpretation (Figure S6 and section Electric field lines topography analysis on Arg171Gln mutant in SI).

The polarisation pronouncedly affects the transferred electron density in the entire binding site. The electrostatic interaction energies (including polarisation) with the 2,4,4'-HBP ligand are −41.4 and −84.0 kcal/mol (for the wild-type Arg171 monomer average and dimer), compared to more favourable −50.3 and −101.6 kcal/mol (for the mutant Gln171 monomer average and dimer). The projected energy difference should correspond to about 1 K difference in thermal shifts, making Arg171Gln an attractive mutation candidate.

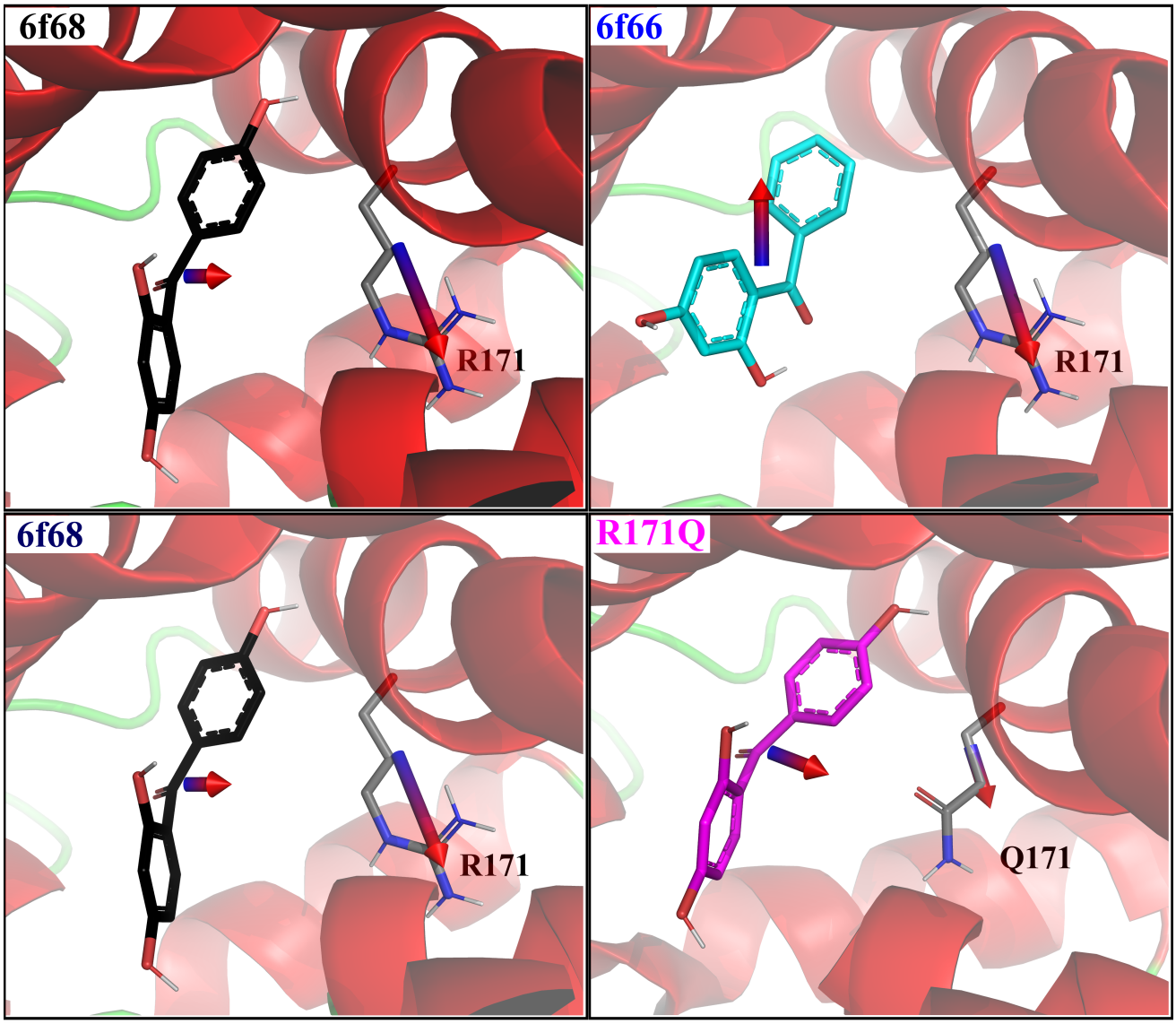


Figure A.7. **Ligand and residue 171 dipole moments.** Strength (arrow length, scaled to 0.21 Å/D) and relative orientation of dipole moments for residue 171 and benzophenone ligands. First row: arginine 171 (on the right of both figures) with a 4'-hydroxy (a, black) and 4'-hydrogen (b, cyan) ligand. Second row: the 4'-hydroxy ligand (on the left of both figures) with wild type arginine 171 in green (c, same as a) and mutated glutamine 171 in light blue (d). We chose glutamine as a mutation candidate for Arg171 that could test the veracity of our computational method. All subfigures depict the binding site A.

A.5 CONCLUSIONS

The methodology presented here and in Nguyen *et al.* (2018) gives results almost identical to the established nEP/MM numerical integration, but in a fraction of time. `Charger` determines electrostatic interaction energies with a precision within the bounds of ±1 kcal/mol for a given charge density model. This remains true even though several newly studied systems contain previously not investigated sulfur atoms. It has to be recalled however that using different charge density models, such as ELMAM2, UBDB and INVARIOM database transfer, results in much larger variations (Bak *et al.*, 2011) and accuracy is estimated about at one quarter in relative value. `Charger`'s purpose is to calculate electrostatic energies, so it does not currently have any features to calculate

other contributions to the total interaction energy. Van der Waals energy term computation is possible with the VMoPro module, using Lennard-Jones parametrizations proposed in the literature.

Charger's *raison d'être* is speed. Its aEP/pMM calculations are 30–80 times faster than VMoPro nEP/MM. They take several seconds for largest examined systems – protein/ligand complexes. This makes Charger an excellent tool to explore crystal or protein electrostatic interaction energies in real time. One can demonstrably improve Charger's speed easily by using appropriate cutoffs, while guarding great accuracy – for example, in a systematic database search where speed becomes a concern.

We have studied the glutathione-benzophenone interaction in detail using Charger. The order of computed Charger protein-ligand energies and experimental thermal shifts is identical in nearly all cases, with high correlation coefficients. The slope of the trend line helps to decide which *in silico* investigations to hand over to the lab.

We have used Charger calculations further to determine contributions of each residue to benzophenone ligand fixation. Charger determines important residues from the electrostatic point of view: Schwartz *et al.* (2018) already suspected most of them from a visual analysis.

Charger also uncovered residues adverse to benzophenone fixation. Particularly interesting was the case of Arg171, which interacts favourably with three benzophenone ligands, but not with the 2,4,4'-HBP. We suspected that mutating Arg171 into a non-charged residue such as a glutamine could remove this electrostatic obstacle to binding. This adds an alternative to strategies that aim to improve or increase the number of existing favourable contacts (Thapa & Raghavachari, 2019).

An *in silico* Arg171Gln mutant meets our expectations. The polarisation correction of the transferred charge density (Leduc, 2019; Leduc *et al.*, 2019) helped to determine the importance of this mutation. If production and purification of the Arg171Gln mutant turns out feasible, thermal shift assays could assess this prediction in the future. Further electrostatic energy studies on numerous protein/ligand systems with more thermodynamic data such as binding constants and enthalpy of binding are needed to assess the importance of polarizing charge densities based on transferred multipolar parameters.

One can easily integrate Charger into different programs. It supports Slater-type basis-functions in general, and Hansen-Coppens charge density models in particular. We release the library code under the permissive Apache 2.0 license to encourage integration into other projects. Our flagship program MoProViewer now comes with a simple and intuitive Charger interface. One can find both the Charger and MoProViewer codes by following this hyperlink:

https://crm2.univ-lorraine.fr/lab/fr/software/mopro/.

Acknowledgements

V.V. is thankful for a PhD fellowship of MENSRT.

A.6 References

Bąk, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B., & Dominiak, P. M. (2011). *Acta Cryst.* A***67***, 141-153.

Bojarowski, S. A., Kumar, P. & Dominiak, P. M. (2016) *ChemPhysChem* **17**.16, 2455–2460.

Buckingham, A. D. (1959). Molecular quadrupole moments. *Chem. Soc. Rev.* **13**, 183–214.

Clementi E. & Roetti C. (1974) *Atomic Data and Nuclear Data Tables* **14**.3, 4–177.

Cimmperman, P., Baranauskienė, L., Jachimovičiūtė, S., Jachno, J., Torresan, J., Michailovienė, V., Matulienė, J., Sereikaitė, J., Bumelis, V. & Matulis, D. (2008) *Biophysics J.* **95**.7, 3222–3231.

Clementi E. & Roetti C. (1974) *Atomic Data and Nuclear Data Tables* **14**.3, 4–177.

Dittrich, B. & Koritsánszky, T. & Luger, P. (2004). *Angewandte Chem. Int. Ed.* **43**.20, 2718–2721.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012) *Acta Cryst.* A**68**, 337–351.

Dominiak, P. M., Volkov, A., Dominiak A. P., Jarzembska, K. N. & Coppens, P. (2009) *Acta Cryst* D**65**, 485–499.

Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007) *J. Chem. Theory Comput.* **3**, 232–247.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Wang, J. Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J. & Kollman, P. (2003) *J. Computat. Chem.* **24,** 1999–2012.

Fournier, B., Bendeif, E. E., Guillot, B., Podjarny, A., Lecomte, C. & Jelsch, C. (2009). *J. Am. Chem. Soc.* **131**.31, 10929–10941.

Guillot, B., Enrique, E., Huder L. & Jelsch, C. (2014) *Acta Cryst.* A**70**, C279.

Hansen, N. K. & Coppens, P. (1978) *Acta Cryst.* A**34**, 909–921.

Hooft, R. W. W., Sander, C. & Vriend, G. (1996) *Proteins* **26**, 363–376.

Huang, L., Massa, L. & Karle, J. (2005) *Biochem.* **44**, 16747–16752.

Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005) *J. Appl. Crystallogr.* **38**, 38–54.

Jones, H. W. (1993) *Int. J. Quantum Chem.* **45**.1, 21–30.

Koga, T., Tatewaki H., & Thakkar, A. J. (1993) *Phys. Rev.* A**47**.5, 4510.

Koga, T., Kanayama K. & Thakkar, A. (2000) *J. Theor. Chim. Acta* **104**.5, 411–413.

Keesom, N. (1921) *Zeitschrift für Physik* **22**.

Kumar, P., Bojarowski, S. A., Jarzembska, K. N., Domagała, S., Vanommeslaeghe, K., MacKerell Jr., A. D. & Dominiak, P. M. (2014) *J. Chem. Theor. Comput.* **10**, 1652–1664.

Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. (2019) *Acta Cryst.* **A75**.2, 398–408.

Leduc, T. (2019) *Vers un potentiel multipolaire quantitatif et transférable aux macromolécules biologiques : une étude méthodologique des effets de la polarisabilité.* PhD in Physics, Université de Lorraine. NNT: 2019LORR0223. tel-02528924.

Leduc, T., Aubert, E., Espinosa, E., Jelsch, C., Iordache, C. & Guillot, B. (2019) *J. Phys. Chem.* A**123**.32, 7156–7170.

Löwdin, P.-O. (1956) *Adv. Phys.* **5**.17, 1–171.

MacLeod, J. M., & Rosei, F. (2014). *Small*, **10**, 1038-1049.

Mandal, S. K., Saha, P., Munshi, P., & Sukumar, N. (2017). *Structural Chemistry*, **28**, 1537-1552.

Massa, L., Keith, T., Cheng, Y., Matta & C. F. (2019) *Chem. Phys. Lett.* **734**, 136650.

Michael, J. R. & Volkov, A. (2015) *Acta Cryst.* A**71**.2, 245–249.

Nguyen, D., Kisiel, Z. & Volkov, A. (2018) *Acta Cryst.* A**74**.5, 524–536.

Nguyen, D., Macchi, P. & Volkov, A. (2020) *Acta Cryst.* **A76**.6, 630–651.

Nguyen, D. & Volkov, A. (2019) *Acta Cryst.* A**75**.3, 448–464.

Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J. Biomol. Screen.* **6**.6, 429–440.

Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0.

Schwartz, M., Perrot, T., Aubert, E., Dumarçay, S., Favier, F., Gérardin, P., Morel-Rouhier, M., Mulliert, G., Saiaf, F., Didierjean, C. & Gelhaye, E. (2018) *Sci. rep.* **8**.1, 1–11.

Thapa, B. & Raghavachari, K. (2019) *J. Chem. Inf. Model.*, **59**.8, 3474–3484.

Volkov, A., Koritsanszky, T. & Coppens, P. (2004) *Chem. Phys. Lett.* **391**, 170–175.

Xing, J., Zhang, R., Jiang, X., Hu, T., Wang, X., Qiao, G., ... & Zheng, M. (2019). *Eur. J. Med. Chem.,* **163**, 281-294.

A.7 Supplementary Information

The following hyperlink leads to the supplementary information of this article:

`https://scripts.iucr.org/cgi-bin/paper?S2059798321008433`

The tables 3.1 and 3.2 and the figure 5.3 have been taken from the Supplementary material and placed in the main body. The following table and figure have been taken from the Supplementary material because they are directly mentioned in the main part of the thesis.

Table A.S6. **Comparison of experimental thermal shifts with calculated interaction energies.** Glutathione transferase thermal shifts with different benzophenone ligands (Table 2), alongside with *Charger* electrostatic interaction energies (precise calculation, non-polarised | polarised).

Ligand	ΔT_d / °C *	E_A / kcal mol^{-1} **	E_B / kcal mol^{-1} **	E_{mean} / kcal mol^{-1}	E_{dimer} / kcal mol^{-1} ***
2,4,4'-trihydroxy-benzophenone	2.87	−20.3 | −36.0	−30.8 | −46.7	−25.6 | −41.4	−52.2 | −84.0
3,4-dihydroxy- benzophenone	4.36	−25.7 | −42.8	−25.9 | −45.1	−25.8 | −44. 0	−55.7 | −87.2
2,4-dihydroxy- benzophenone	4.96	−37.6 | −56.7	−46.1 | −69.1	−41.8 | −62.9	−83.2 | −125.4
2,3,4-trihydroxy-benzophenone	5.69	−35.0 | −55.5	−49.4 | −75.4	−42.2 | −65.5	−82.4 | −127.3

* The relative scale uses apo-GST denaturation temperature as zero.
** Interaction energy between a GST monomer and the ligand bound in its active site.
*** Interaction energy between the GST dimer and both ligands.

Correlation coefficients (non-polarised energies):

(%)	E_A	E_B	E_{mean}	E_{dimer}
Pearson (E with ΔT_d)	−96	−69	−83	−87
Spearmann (E with rank ΔT_d)	98	78	89	92
Spearmann (rank E with rank ΔT_d)	−100	−60	−80	−100

Correlation coefficients (polarised energies):

(%)	E_A	E_B	E_{mean}	E_{dimer}
Pearson (E with ΔT_d)	−95	−75	−84	−84
Spearmann (E with rank ΔT_d)	96	81	89	90
Spearmann (rank E with rank ΔT_d)	−100	−60	−80	−80

Note: the minus sign in correlation coefficients means that the corresponding trend line has a downward slope.

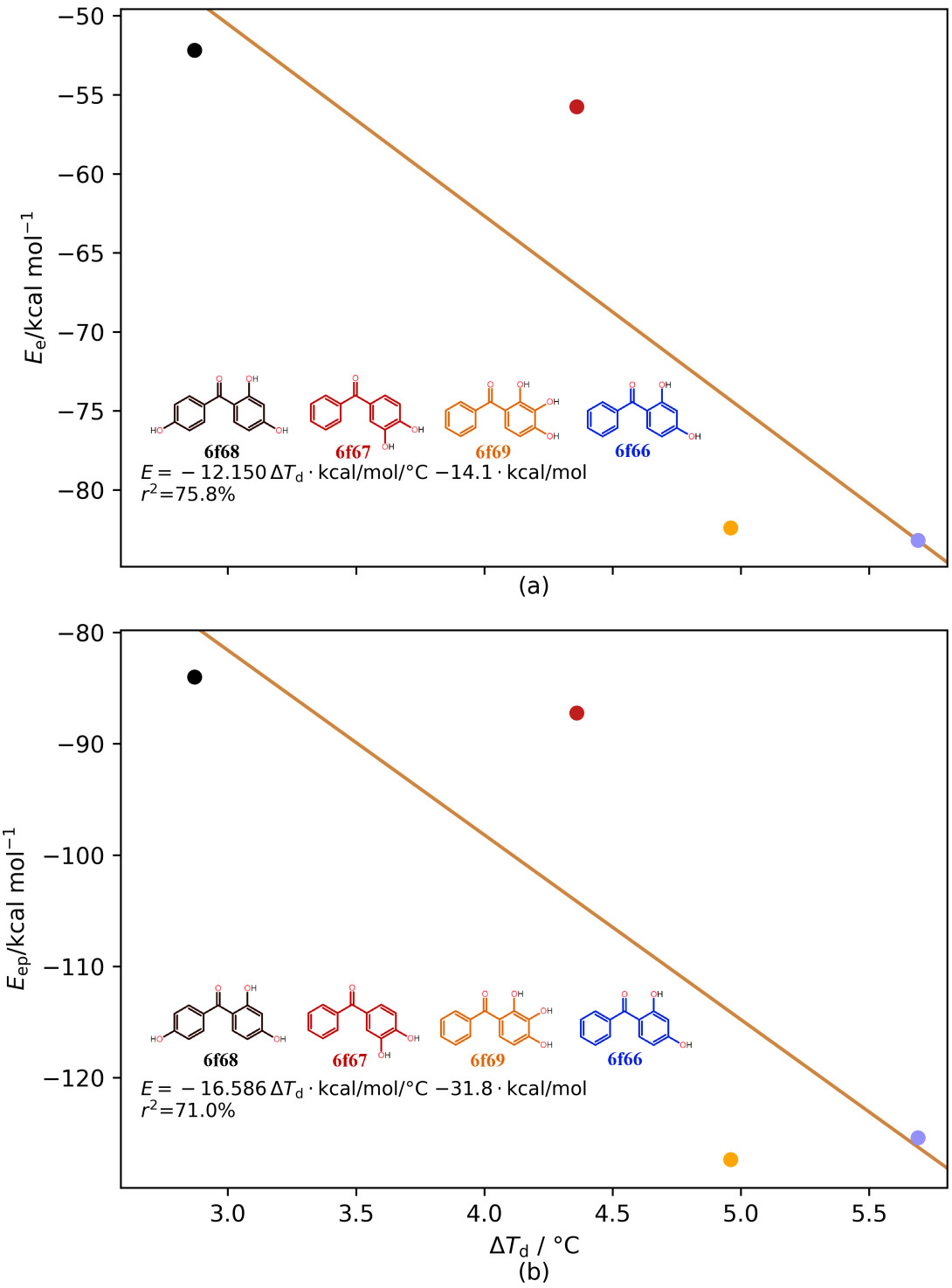


Figure A.S2. **Relationship between experimental thermal shifts and computed energies for GST dimers.** Relationship between computed electrostatic interaction energies (a: non-polarised, b: polarised) from Charger calculations and experimental denaturation temperatures for glutathione transferase and four ligands. Energies shown are the interactions between the whole protein (a dimer) and two ligands in their binding sites. Values given in table A.S6.

B An unusual intermolecular interaction between a lone pair and an electron-rich π-electron system of a quinoid dianion

Article published in the *Crystal Growth & Design* journal: https://pubs.acs.org/doi/10.1021/acs.cgd.1c00492.

Vedran Vuković,[1] Tomislav Piteša[2], Christian Jelsch,[1] Emmanuel Wenger,[1] Krešimir Molčanov*[2]

[1] CRM2, CNRS, Université de Lorraine, 54000 Nancy, France
[2] Ruđer Bošković Institute, Bijenička 54, HR-10000 Zagreb, Croatia
* e-mail: kmolcano@irb.hr

B.1 Abstract

The crystal structure of lithium 2,5-dihydroxyquinonate dihydrate bis(dimethylsulfoxide) solvate [$Li_2DHQ(H_2O)_2(DMSO)_2$] reveals a close contact between the lone pair on the electropositive S atom of DMSO and the dianionic quinoid ring. This contact, resembling a π-hole interaction involving a dianion as an acceptor was studied in detail by a combination of X-ray charge density and quantum chemical modelling. Topology of electron density reveals a bond path and a (3, −1) critical point between the S atom of the DMSO and a C atom of the quinoid ring, while analysis of electrostatic potential indicates that the interaction between these two atoms is repulsive. However, the global electrostatic interaction between the DMSO and the dihydroxyquinonate moieties is attractive (−19.1 kJ/mol), as well as total *in vacuo* interaction energy (−42.8 kJ/mol).

Keywords π-hole interactions, 2,5-dihydroxyquinone, charge density, crystal structure, DFT computations

B.2 Introduction

A class of intermolecular interactions, which recently came into a research focus, are those involving electron density depletions of a π-system, generally named π-hole interactions.[1-4] They typically involve electrostatic attraction of an electron-rich group with an electron-depleted region of a π-electron system.[3-6] Donation of a lone electron pair (lp) into an empty π* orbital of the π-system (i.e. lone pair···π or $n\rightarrow\pi^*$ interactions) explains some cases of π-hole interactions.[4,7-9] Anions also form anion···π interactions with electron-depleted conjugated rings.[2,10-14] Formation of anion···π, lp···π and $n\rightarrow\pi^*$ interaction necessitates the presence of a positive electrostatic potential (ESP) zone in a conjugated system (i.e. a π-hole) on the acceptor. Commonly used acceptors include aromatic rings with multiple electron-withdrawing substituents; non-aromatic perhalogenoquinones have been investigated recently.[15-17] Generally, quinoid rings are inherently electron-poor due to the presence of strongly electron-withdrawing carbonyl groups. Their influence is exacerbated by the π electrons which are much more localised than in the aromatics (*i.e.* single and double bonds in quinones can be easily distinguished[18]). Electronegative substituents such as halogens further deplete the electron density of quinoid rings, resulting

in very pronounced π-holes located at carbonyl C atoms.[15,19-21] Intermolecular contacts shorter than the sum of van der Waals radii were observed between perhalogenquinones and halide anions.[15-17]

So far, only severely electron-depleted π systems were considered as possible acceptors of π-hole interactions. In this paper, we present a detailed study of a serendipitously obtained compound with a contact resembling a lone pair···π interaction involving an electron-rich dianionic quionid ring 2,5-dihydroxyquinonate (**DHQ**). To reveal more details on the nature of this counterintuitive interaction, we have studied it by a combination of experimental X-ray charge density and quantum chemical modelling.

B.3 Results and Discussion

The studied compound, lithium 2,5-dihydroxyquinonate dihydrate bis(dimethylsulfoxide) solvate [$Li_2DHQ(H_2O)_2(DMSO)_2$, **1**, Scheme B.1] was serendipitously obtained by recrystallization of lithium 2,5-dihydroxyquinonate[22] in dimethylsulfoxide (DMSO). Asymmetric unit of the compound comprises a half of a **DHQ** dianion, a lithium cation, a water molecule and a molecule of DMSO.

Structure of the 2,5-DHQ dianion: induction effect of substituents quantified by analysis of charge density

Bond lengths in the **DHQ** dianion (Table S3) is in a good agreement with its other alkali salts,[23] confirming the structure shown in Scheme B.1. Several substituted 2,5-dihydroxyquinones also have two delocalised π systems separated by two single C-C bonds: 3,6-dichloro analogue (chloranilate dianion, **CA**),[18,24-26] 3,6-dibromo analogue (bromanilate dianion, **BA**),[27] 3,6-dinitro analogue (nitranilate dianion, **NA**)[28,29] and 3-cyano-6-chloro analogue.[30]

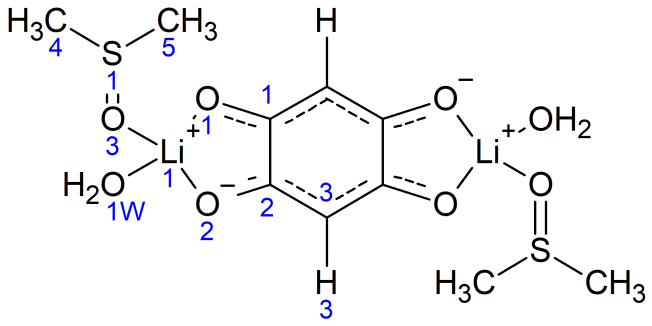
Experimentally determined charge density for chloranilate[18] and nitranilate dianions[28] can be compared to **DHQ** to quantify substituent inductive effect. Electron-withdrawing substituents enhance negative charge delocalisation and thus increase hydroxyl group acidity.[24,30] Their effect on molecular geometry is negligible. Bond orders calculated from electron density show a clear trend (Table B.1): electron density in the carbon ring is reduced as electron-withdrawing power of the substituents increases. This effect is the most pronounced for the delocalised bonds C2-C3 and C3-C1, whose estimated bond order decreases from 1.43 in **DHQ** to 1.25 in **CA**[18] and to less than 1.0 in **NA**.[28] Order of the formally single bond C1-C2 decreases from 0.77 in **DHQ** to 0.67 in **NA**.[28] Electron-withdrawing substituents seem not to have an effect on C-O bonds. ESP maps also reveal the decrease of electron density in the carbon skeleton (Fig. B.1), while Table S2 shows an increase of negative charges on substituents.

The distribution of charges on dihydroxyquinonate ion (Tables S1 and S2) is comparable to that of chloranilate[18,31] and the nitranilate anion.[28] The oxygen atoms are more negative, while the hydrogen is more positive than comparable substituents in other quinones. It appears as if the electron density from these hydrogen atoms re-locates to the oxygen atoms on the edges of π-systems.

Interaction between the lone pair of DMSO and the π-system of the DHQ dianion

The basic building block of the crystal structure of **1** is a centrosymmetric moiety $Li_2DHQ(H_2O)_2(DMSO)_2$, with the **DHQ** ring centroid located at the crystallographic inversion centre (Scheme B.1, Figure S6). The lithium cation has a tetrahedral coordination by O1 and O2 atoms from **DHQ**, water molecule oxygen O4, and DMSO oxygen O3. There are two symmetry-independent O-H···O hydrogen bonds and seven weaker C-H···O, which form layers parallel to $(\bar{1}01)$ (Fig. B.2) containing alternatively (**DHQ**, H_2O, Li^+) and DMSO moieties. Therefore, there is no π-stacking of **DHQ** rings in the crystal structure.

A close contact between the DMSO sulphur atom and the **DHQ** ring stands apart as the most interesting feature in the crystal structure (Fig. B.3). The closest separation is S1···C2 of 3.2481(7) Å,



Scheme B.1. Lewis structural formula and atom numeration scheme of $Li_2DHQ(H_2O)_2(DMSO)_2$ (**1**).

Table B.1. Comparison of bond orders in the quinoid dianions **DHQ**, **CA**[18] and **NA**.[28]

Bond	**DHQ**	**CA**	**NA**
C1-C2	0.77	0.76	0.67
C2-C3	1.45	1.25	0.90
C3-C1^i	1.41	1.25	0.76
C1-O1	1.34	1.48	1.47
C2-O2	1.30	1.45	1.37

Symmetry operator: *i*) $-x, -y, -z$.

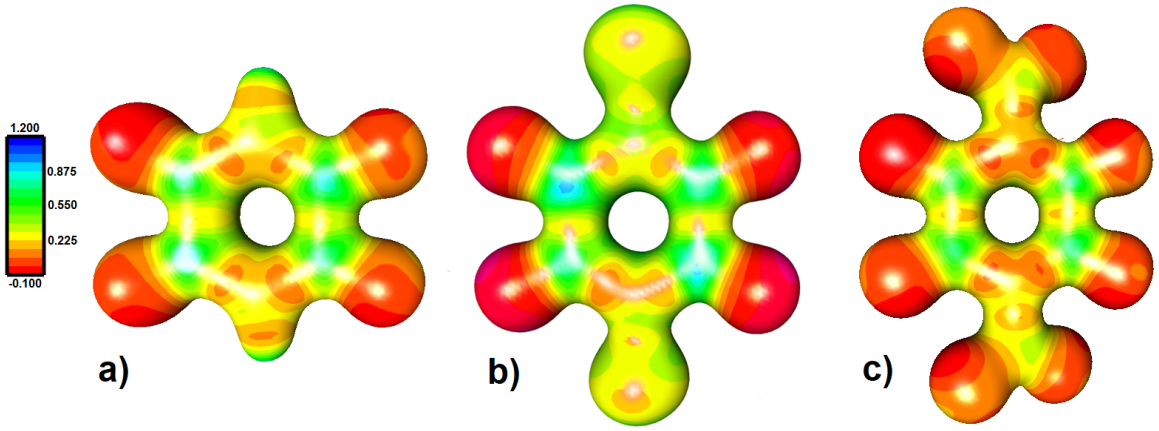


Figure B.1. Comparison of electrostatic potentials plotted onto an electron density isosurface of 0.5 e $Å^{-3}$ generated by the **DHQ**, **CA**[18] and **NA**[28] quinoid dianions.

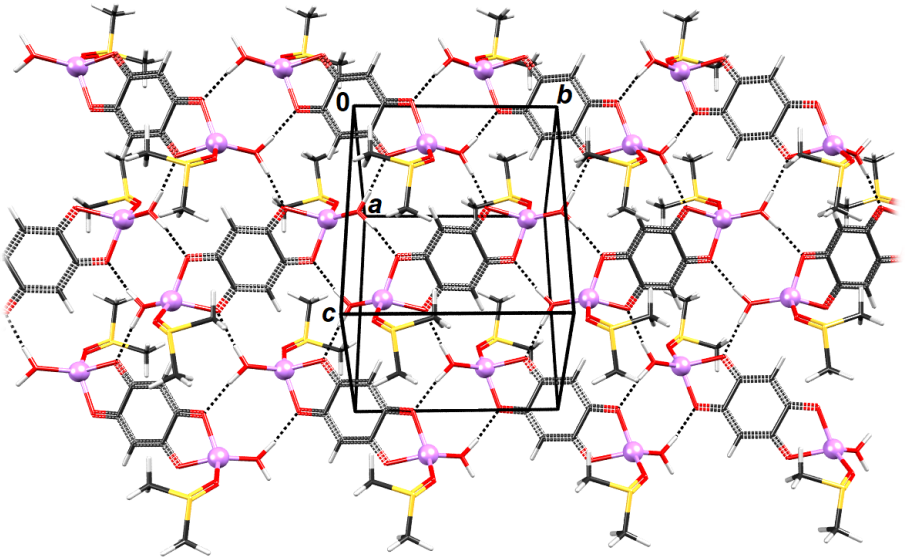


Figure B.2. Crystallographic autostereogram of **I** viewed approximately normal to the $(\bar{1}01)$ plane. The H-O-H···O-C hydrogen bonds are shown as dotted lines.

followed by S1···C1 of 3.4046(7) Å; the distance between S1 and the centre of gravity of **DHQ** ring is 3.461 Å. The molecules arrange so that the lone electron pair of S1 points towards the π-system of the quinoid ring (Fig. B.3). The interaction between a **DHQ** dianion and a lone pair is not expected to be electrostatically attractive: electrostatic potential at atoms C1 and C2 is higher than the rest of the ring, but still negative (Figs. B.1 and B.4).

The ESP generated by DHQ on the Hirshfeld surface of the "π-hole" (at atom C2) is −0.55 V, while the DMSO ESP at the position of the sulphur lone pair it is −0.05 V. Despite the local electron-depletion of atom C2, the ESP in its surroundings is negative, due to the global charge of **DHQ** dianion potential. Moreover, the q_{model} atomic charge derived from P_{val} is +0.41 e for S1, while on the **DHQ** side it is 0.00 e for C1 and +0.01 e for C2, meaning that these atoms atoms are richer in electrons than the S1 sulphur (Fig. B.5). From an electrostatic point of view, this interaction should not be attractive.

The new *Charger* electrostatic interaction energy tool[32] expands this simple charge-based analysis. It considers the full charge density description around selected pseudo-atoms. Both S1···C1 and S1···C2 electrostatic interaction energies are slightly positive (+0.9 and +2.5 kJ/mol, respectively): the pseudo-atoms in direct contact repulse each other electrostatically. However, the interaction becomes attractive when the S1=O3 group is considered instead of just the S1 atom: respective electrostatic energies of the contacts (O3=S1)···C1 and (O3=S1)···C2 become −0.5 and −1.2 kJ/mol. Presumably the electrostatic repulsion between S and C is compensated by the electronegative oxygen. This occurs often with the multipolar model, in which the ensemble of the two pseudo-atoms that form a covalent bond describe its density. However, the electrostatic energies between S1 and the two C3 atoms in the quinoid ring, with respective distances of 3.6386(7) and 3.8371(7) Å are −36.2 and −32.8 kJ/mol. This attractive energy is mostly contributed by the deformation energy −37.9 and −35.7 kJ/mol (0.03 e $\mathrm{Å}^{-3}$), which originates from the $N_{\mathrm{val}} - P_{\mathrm{val}}$ charges S1 (+0.41e) and C3 (−0.24e). The overall **DHQ**···DMSO interaction energy (−19.1 kJ/mol) indicates a strong global electrostatic attraction between them.

The topology of electron density revealed multiple bonding critical points (cp, Fig. B.6, Table B.2). The strongest ones correspond to O-H···O hydrogen bonds, followed by S1···C2 contact (electron density at the cp is 0.056 e $\mathrm{Å}^{-3}$), while C-H···O hydrogen bonds are considerably weaker. However, the presence of (3, −1) critical points does not always imply the existence of

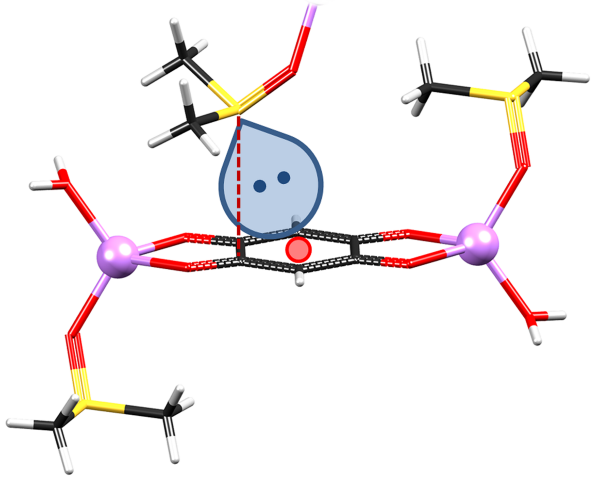


Figure B.3. Close contact between DMSO and **DHQ** dianion. The lone pair of the S1 atom (shown in blue) is directed approximately towards the ring centroid (shown as red sphere); contact S···C is shown as red dashed line.

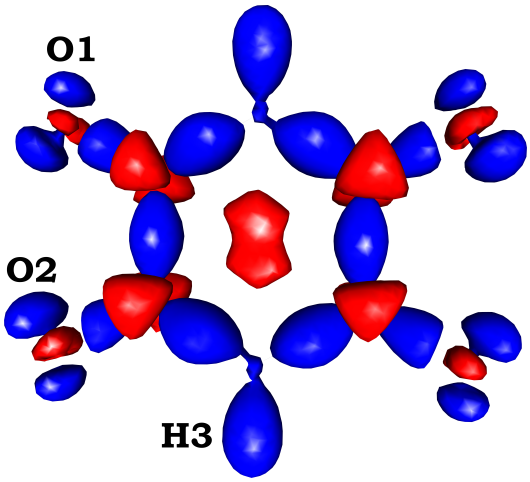


Figure B.4. Deformation electron density map of the **DHQ** dianion. Contours: blue: +0.25 e/Å^3, red : −0.125 e/Å^3.

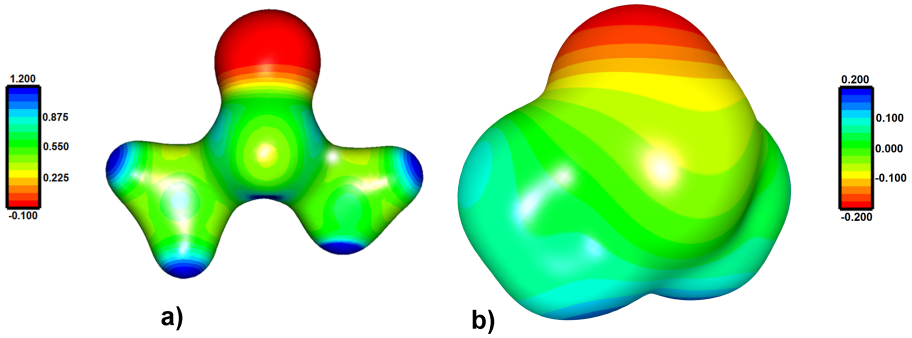


Figure B.5. Electrostatic potential of a DMSO molecule plotted onto an electron density isosurface of a) 0.5 e Å^{-3} and b) 0.005 e Å^{-3}.

attractive interactions,[33,34] especially in the case of weak intermolecular interactions.[35] The S1···C2 atom contact is not attractive from a pure electrostatic point of view, but is the result of the global attraction between the **DHQ** and DMSO entities. However, quantum-chemical *in vacuo* computations, employing density functional theory, tell a different story. First, geometry optimization of DHQ^{2-}···DMSO complex *in vacuo*, starting from the geometry as found in crystal structure of **1**, did not yield a complex with either S···C or S···O interaction, but rather a C-H···O hydrogen-bonded complex. However, optimization of the $Li_2DHQ(H_2O)_2(DMSO)_2$ moiety yielded a minimum with a little bit distorted S···ring interaction (Fig. S9, Table S6). This is another proof that this interaction is in fact just a consequence of all other (stronger) interactions.

Nevertheless, total (electronic, not electrostatic) energy of interaction between **DHQ** and DMSO at the crystal-structure geometry of **DHQ**···DMSO appears strongly attractive (−42.8 kJ/mol). To investigate further the nature of this interaction, we analysed the electron density difference (EDD) upon formation of the interaction (Fig. B.7). This analysis show that the interaction shifts electron density from the lone-pair region of S1 closer to its nucleus and to the O1 of **DHQ**. The electron depletion of the sulphur atom S1 resulting from the electron withdrawing O=S oxygen atom (Fig. 4b) is likely a result of electron repulsion caused by the close contact with the dianion.

B.4 Conclusions

The $Li_2DHQ(H_2O)_2(DMSO)_2$ (**1**) crystal structure comprises an unusual intermolecular contact resembling a π-hole interaction involving a dianion as an acceptor. A detailed study of the close contact between the sulphur atom of DMSO and the **DHQ** dianion by a combination of X-ray charge density and quantum chemical modelling was undertaken to study nature of this unusual interaction.

The topology analysis indicates the presence of a critical point, but the analysis of the ESP tells that the interaction is not localised, and the electrostatic interaction between pseudoatoms S1 and C2 is in fact slightly repulsive. On the contrary, when taking into account the entire DMSO molecule and a **DHQ** dianion, the electrostatic interaction between these two moieties becomes attractive, the total estimated energy being −19.1 kJ/mol. DFT *in vacuo* computations and analysis of ESP indicate strong electrostatic attraction between $S1^{\delta+}$ and $C3^{\delta-}$ atoms. Therefore, we conclude that there is no local attraction between S1 and C2 atoms (despite presence of a bonding cp), but that the total interaction between a **DHQ** dianion and a molecule of DMSO is attractive. Its main component is electrostatic. Thus, presence of a bonding cp between S1 and C2 is a result of a generally attractive interaction between **DHQ** and DMSO as well as a combination of other interactions present in the crystal field.

This work shows that a short intermolecular contact *and* a presence of a bonding cp does not always characterise an attractive interaction. Therefore, a care should be taken when interpreting novel and apparently counterintuitive interactions.

In addition, we have quantified induction effect of substituents on electron density in the quinoid ring. A comparison of electron density in **DHQ**, **CA**[18] and **NA**[28] dianions revealed a different distribution of ESP. Carbon-carbon bond orders in the quinoid ring decrease as electron-withdrawing capability of substituents increases.

B.5 Experimental

Preparation

Diaqua*bis*(dimethylsulfoxyde)-μ-(2,5-dihydroxyquinonato)dilithium ($[Li_2\{\mu\text{-}(C_6H_2O_4)\}(H_2O)_2\{(CH_3)_2SO\}_2]$ (**I**), Scheme B.1) crystals synthesis involves two steps. The first entails the neutralisation of a 2,5-dihydroxyquinone solution with an equimolar amount of lithium carbonate, forming lithium 2,5-dihydroxyquinonate, as described previously[23,36]. 2,5-dihydroxyquinonate crystalizes by evaporation of this solution in air. The crystals were not large enough for a high-resolution diffraction experiment. They were then re-crystallised from a dimethylsulfoxyde solution. Well-developed crystals of $[[Li_2\{\mu\text{-}(C_6H_2O_4)\}(H_2O)_2\{(CH_3)_2SO\}_2]]$ (**I**) grew

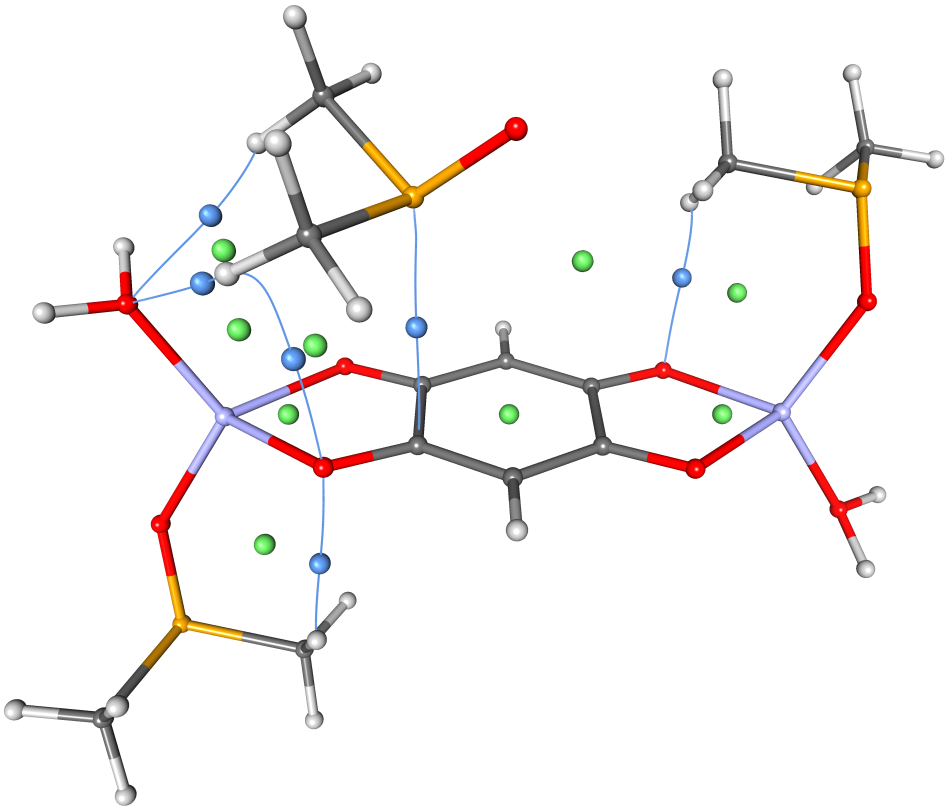


Figure B.6. Critical points in a $Li_2(DHQ)(H_2O)_2DMSO(2)$. DMSO moiety. (3,−1) cps are shown as blue spheres, (3, +1) ring cps as green spheres and bond paths are shown as blue lines.

Table B.2. Total electron density ρ_{tot} and Laplacian $\nabla^2\rho$ at the intermolecular contacts (critical points of the (3, −1) type between molecules) in **I**. Experimentally determined cps are printed regular and theoretical ones are in italic.

$A\cdots B$	d / Å	ρ_{tot} / e Å^{-3}		$\nabla^2\rho$ / e Å^{-5}		G_{cp} / e Å^{-3}		V_{cp} / e Å^{-3}		$\lvert V_{cp}/G_{cp}\rvert$	
		exp.	theor.	exp.	theor.	exp.	theor.	exp.	theor.	exp.	theor.
S1···C2^{i}	3.248(7)	0.056	*0.063*	0.68	*0.76*	0.006	*0.006*	−0.004	*−0.005*	0.67	*0.83*
H1WA···O1^{ii}	1.766(6)	0.308	*0.231*	2.34	*1.55*	0.033	*0.021*	−0.042	*−0.026*	1.27	*1.24*
H1WB···O2^{iii}	1.739(6)	0.337	*0.258*	2.41	*1.45*	0.036	*0.022*	−0.047	*−0.030*	1.31	*1.36*
H3···O3^{iv}	2.963(7)	0.013	*0.018*	0.24	*0.25*	0.002	*0.002*	−0.001	*−0.001*	0.50	*0.50*
H4A···O1W^{v}	2.859(8)	0.020	*0.028*	0.33	*0.31*	0.002	*0.002*	−0.001	*−0.002*	0.50	*1.00*
H4B···O3^{vi}	2.464(6)	0.043	*0.072*	0.84	*0.60*	0.006	*0.006*	−0.004	*−0.005*	0.67	*0.83*
H4C···O1^{iii}	2.674(6)	0.031	*0.040*	0.51	*0.43*	0.004	*0.004*	−0.002	*−0.003*	0.50	*0.75*
H5A···O1W^{v}	2.649(8)	0.029	*0.045*	0.54	*0.48*	0.004	*0.004*	−0.002	*−0.003*	0.50	*0.75*
H5B···O2^{vii}	2.524(9)	0.050	*0.060*	0.79	*0.74*	0.006	*0.01*	−0.005	*−0.01*	0.83	*0.83*
H5C···O3^{viii}	2.671(10)	0.037	*0.033*	0.56	*0.45*	0.004	*0.004*	−0.003	*−0.002*	0.75	*0.50*

Symmetry operation on B: (*i*) $^3/_2 - x$, $^1/_2 + y$, $^1/_2 - z$; (*ii*) $1 - x$, $2 - y$, $1 - z$; (*iii*) $-^1/_2 - x$, $^3/_2 - y$, $-^1/_2 - z$; (*iv*) $^1/_2 - x$, $^3/_2 - y$, $^1/_2 - z$; (*v*) $^1/_2 - x$, $^3/_2 - y$, $-^1/_2 - z$; (*vi*) $1 - x$, $2 - y$, $-z$; (*vii*) $1 - x$, $1 - y$, $1 - z$; (*viii*) $^3/_2 - x$, $-^1/_2 - y$, $^1/_2 - z$.

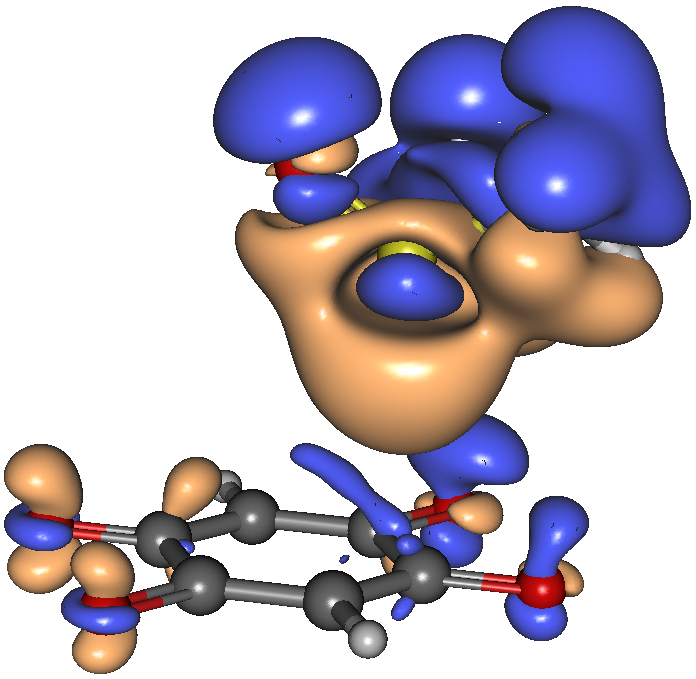


Figure B.7. Isosurface of theoretical electron density difference (EDD) upon formation of $DHQ^{2-}\cdots$DMSO interaction, $\delta\rho = \rho(DHQ^{2-}\cdots DMSO) - \rho(DHQ^{2-}) - \rho(DMSO)$, for isovalue ±0.0007 a.u. (brown – negative isovalue, blue – positive isovalue).

over the course of six months at room temperature.

X-ray diffraction and multipolar refinement

Collection of diffraction data. The diffraction data for a single crystal of **1** were collected using a Bruker D8 Venture diffractometer equipped with a molybdenum micro-focus X-ray tube with a mirror monochromator, and a PHOTON 100 CMOS detector. MoK_α radiation (λ = 0.71073 Å) was used. `Bruker SAINT` software performed data reduction and numerical absorption correction based on crystal shape.[37] The full dataset contains 97590 reflections up to the resolution of $\sin(\theta)/\lambda$ = 1.0 $Å^{-1}$, 6889 of which are symmetry-independent. The data set is 99.7% complete, with R_{int} = 7.06%; detailed statistics can be found in Table B.3.

Structure solution and initial spherical refinement. The `Olex2` graphical user interface served to solve and refine the structure of **1**.[38] To this end, it employed exterior programs `SHELXS97`[39] and `SHELXL-2017`.[40] The structures were refined using the least-squares method. Anisotropic displacement parameters were used for all atoms, except hydrogens. Hydrogen atoms were found from difference Fourier maps and refined without restraints. Molecular graphics were made using `CCDC-Mercury`[41] and `POV-Ray`.[42]

Charge density refinement. The program `Import2MoPro` from the `MoProSuite` transformed the solved and refined crystal structure of **1** into the appropriate format for further refinement.[43] `SORTAV` removed symmetry-forbidden reflections and averaged the remaining ones.[44]

The program `MoPro` refined the crystal structure using the Hansen-Coppens multipolar model.[43] A gradual refinement of following parameters was executed: scaling factor coefficients, atomic coordinates, anisotropic displacement parameters, valence populations, multipolar coefficients (none for lithium atoms; up to dipolar level for hydrogen atoms; octupolar levels for carbon and oxygen atoms; hexadecapolar level for the sulphur atom), and expansion/contraction parameters κ and κ'. The least-squares refinement had the goal of min-

Table B.3. Details and statistics about gathered crystallographic data, as well as structure and charge density refinement of **1**.

Compound	**1**
Empirical formula	$C_5H_6LiSO_4$
Empirical mass / (g/mol)	209.56
Crystal dimensions / mm^3	451 x 336 x 132
Space group	$P\,2_1/n$
a / Å	8.1521(7)
b / Å	8.9827(7)
c / Å	11.1947(9)
α / °	90
β / °	93.238(4)
γ / °	90
Z	4
V / $Å^3$	818.46(12)
D_{calc} / (g/cm^3)	1.397
μ / mm^{-1}	0.355
θ range / °	2.91 – 45.42
T / K	100(2)
λ / Å	0.71073
Diffractometer type	Bruker D8 Venture
h, k, l range	$-16 < h < 15$
	$-17 < k < 17$
	$-22 < l < 22$
Resolution / Å	0.500
Reciprocal resolution / $Å^{-1}$	1.000
Gathered reflections	97 590
Independent reflections	6889
Reflections with $I \geq 2\sigma$	4871
Absorption correction	Multi-scan
T_{min}	0.856
T_{max}	0.955
Statistical weighting scheme	$\{1.06\sigma^2(F_{obs})\}^{-1}$
Completeness (%)	99.7
R_{int} (%)	7.06
$R(F)$ (%)	5.007
$R(F^2)$ (%)	3.307
Goodness-of-fit	1.003
Number of parameters	304
Number of constraints/restraints	143
$\Delta\rho_{min}$ / e $Å^{-3}$	−0.44
$\Delta\rho_{max}$ / e $Å^{-3}$	0.46

imising the difference between observed (I_{obs}) and calculated intensities (I_{calc}).

A calculation on the SHADE3 server determined anisotropic displacement parameters of hydrogen atoms from 2,5-dihydroxyquinonate dianion and dimethylsulfoxyde.[45,46] Calculated values of U_{ani}(H) were constrained to those from SHADE3 during the refinement. There was insufficient data to determine U_{ani}(H) for the water molecule because the program SHADE3 requires at least four non-hydrogen atoms be present in the molecule.

Restraints on X–H bond lengths were as follows: 0.983(2) Å for O–H bonds in water, 1.083(2) Å for C–H bonds in 2,5-dihydroxyquinonate ion and 1.077(2) Å for C–H bonds in methyl groups of DMSO.[47] Furthermore, dipolar, quadrupolar and expansion/contraction parameters for the water molecule were constrained to those from the ELMAM2 database.[48]

Several $R_{F,free}$ and $wR^2_{F,free}$ calculations with 20 refinement cycles determined optimal restraint values for multipolar local symmetry and valence/multipolar population similarities. The minimum of $R_{F,free}$ and $wR^2_{F,free}$ curves was the basis to determine optimal restraint values.

We used three graphical techniques to evaluate model quality: the DR plot (values of $(I_{obs} - I_{calc})/\sigma(I_{obs})$ *vs.* theoretical quantiles of a standard normal distribution; the points should be on the $y = x$ line),[49,50] the resolution plot (values of I_{obs}/I_{calc} *vs.* reciprocal resolution $\sin(\theta)/\lambda$; the points should be on the $y = 1$ line),[50] and the fractal dimension plot (value of residual density at different iso-surfaces *vs.* fractal dimension of those iso-surfaces; the points should form a parabola with the maximum at (0, 3) in case of purely Gaussian noise). The first two plots were made with XDRKplots, and the third with jnk2RDA.[51,52]

The ground to decide in which directions to proceed with the refinement was the compromise between improving these plots, residual density global minimum/maximum values and deformation density contour plots. In the final cycles of refinement, the resolution of diffraction data used for refinement was limited to $\sin(\theta)/\lambda = 0.95$ Å^{-1} due to I_{obs}/I_{calc} scaling deviating to values higher than unity at higher resolution (Fig. S1).

Charge density analysis. VMoPro calculated total charge density cubes for the entire complex (two asymmetric units related by inversion in the centre of the quinone ring). VMoPro further calculated ESPs, bond critical points and their parameters (density, eccentricity, density Laplacian *etc.*).[43]

Atomic charges were calculated in two ways. The first is directly from the Hansen-Coppens model, which contains the parameter of valence population P_{val}. This charge is hence called the model charge, calculated as the difference between neutral and modelled valence populations: $q_{model} = N_{val} - P_{val}$. The second way to calculate charges is an integration of charge density in the volume of the space designated to each atom. The program Critic2 calculated topological charges using the Bader partitioning method (q_{Bader}) from total charge density cubes.[53]

Topological bond orders were determined using an empirical formula:

$$n_{topo} = a + b\lambda_1 + c(\lambda_2 + \lambda_3) + f\rho_{cp}. \qquad \text{(B.1)}$$

The values *a*, *b*, *c* and *f* are empirical parameters taken from literature sources,[54-56] ρ_{cp} represents the charge density at the critical point, and λ_1, λ_2, λ_3 represent density Hessian values at the critical point ($\lambda_1 \geq \lambda_2 \geq \lambda_3$).

MoProViewer created visual representations of molecules with charge density surfaces.[57] VMoPro generated most two-dimensional property plots. Ffor standard deviation plots, MoProViewer 2D slices from 3D cubes functionality were applied.[57]

The calculation of standard deviations for all these parameters required several tools. MoPro can generate structures with errors added to all parameters (from the atomic coordinates to the multipolar coefficients).[58] These errors are taken from the standard normal distribution and transformed to random deviations appropriate for each individual parameter (coming from the variance-covariance matrix during crystallographic refinement).

Twenty structures with errors generated in this way enable calculations of standard deviations for model parameters and derived properties. All these structures were processed as described above (charge density 3D cubes calculation, critical point analysis, topological charge

integration, topological bond order determination). Three-dimensional cubes of charge density standard deviations were calculated in the "3D map operations" module of MoProViewer.[57] Critical point parameters and topological charges were loaded into a Jupyter-Notebook-enabled[59] Python script[60] which calculated their standard deviations using the Pandas module.[61] Additionally, it estimated topological bond orders according to the formula above, and their standard deviations.

Quantum chemical computations

Atomic coordinates after multipolar refinement were converted with MoPro into an input file for the quantum chemistry program Crystal14[62,63]. This program was used to calculate periodical wave-function without prior geometry optimisation of initial crystal structures. Periodic DFT using the B3LYP functional was used to determine the adequate wave-function which describes the crystal. Basis sets of functions for carbon,[64] hydrogen,[64] oxygen[64] and sulphur[65] atoms were of 6-31G(d,p) type, adapted for periodical calculations according to recommendations from literature. The adaptations were not found for lithium atoms, so the 6-31G(d,p) non-periodic coefficients were used.[66]

The amount of available memory for integral calculation was increased to ensure their correctness. Tolerance boundaries for two-electron integrals during orbital pseudo-overlap were increased according to recommendations from literature.[65] Self-consistent bond-perturbed Kohn-Sham matrix calculation was interrupted after 100 cycles maximum. Kohn-Sham matrix eigenvalues are reduced by 0.6 hartree to improve wave-function convergence.[63]

The program properties from the Crystal14 program package derived structural factors up to $\sin(\theta)/\lambda = 1.28$ Å^{-1} resolution from the calculated wave-function.[62,63] Calculated structure factors were used with experimentally-determined atomic coordinates to refine a Hansen-Coppens charge density using MoPro. The following parameters were gradually refined: valence populations, multipolar coefficients (like in the experimental model, except for hydrogen atoms where quadrupolar parameters were also refined) and the values of parameters κ and κ'. Charge densities were refined with the least-squares method, by minimising the difference of observed (F_{obs}) and calculated structural factors (F_{calc}).

Physically non-justifiable κ and κ' values in experimentally determined charge densities were restrained to parameters obtained from this theoretical model. Lithium and sulphur atoms had problems with κ values, so the refinement had them restrained to 0.997(10) and 1.002(10) respectively. Similarly, κ' values of oxygen atoms had to be restrained to 0.907(10) for O1, 0.999(10) for O2 and 1.098(10) for O3.

In vacuo DFT calculations were performed with the Gaussian 09 program.[67] Geometry optimizations were performed on the B3-LYP[68] /jun-cc-pVTZ[69] level of theory with Grimme's D3 dispersion correction,[70] starting from the crystal-structure geometries of DHQ···DMSO and $Li_2DHQ(H_2O)_2(DMSO)_2$ moieties. The calculation of interaction energy was performed on the crystal-structure geometry of DHQ^{2-}···DMSO complex on the same level of theory and with counterpoise scheme of Boys and Bernardi to account for the BSSE.[71] The EDD calculation was performed on the CAM-B3LYP[72]/jun-cc-pVTZ level on the crystal-structure geometry of DHQ^{2-}···DMSO.

B.6 Supporting Information

Additional details on refinement of an electron density; molecular geometry; atomic charges and electrostatic potential. This material is available free of charge via the Internet at https://pubs.acs.org/doi/10.1021/acs.cgd.1c00492.

Author Information

Corresponding Author. Krešimir Molčanov, Ruđer Bošković Institute, Bijenička 54, HR-10000 Zagreb, Croatia, e-mail kmolcano@irb.hr

Author Contributions. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources. The Croatian Science Foundation financed this work, via its grant no. IP-2019-04-4674. V.V. is thankful for the French PhD fellowship of MENSRT.

Acknowledgements

We thank the X-ray diffraction facility PMD^{2}X of the Institut Jean Barriol, Université de Lorraine, for X-ray diffraction measurements. (http://crm2.univ-lorraine.fr/lab/fr/services/pmd2x)

Abbreviations

DHQ, 2,5-dihydroxyquinonate dianion; CA, chloranilate dianion; BA, bromanilate dianion; NA, nitranilate dianion; DMSO, dimethylsulfoxide; EDD, electron density difference; cp critical point; ESP, electrostatic potential

B.7 References

[1] Quiñonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M. Anion–π Interactions: Do They Exist? *Angew. Chem. Int. Ed.* **2002**, 41, 3389–3392.

[2] Frontera, A.; Gamez, P.; Mascal, M.; Mooibroek, T. J.; Reedijk, J. Putting Anion–π Interactions Into Perspective. *Angew. Chem. Int. Ed.* **2011**, 50, 9564–9583.

[3] Kozuch, S. Should "anion–π interactions" be called "anion–σ interactions"? A revision of the origin of some hole-bonds and their nomenclature. *Phys. Chem. Chem. Phys.* **2016**, 18, 30366–30369.

[4] Angarov, V.; Kozuch, S. On the σ, π and δ hole interactions: a molecular orbital overview. *New. J. Chem.* **2018**, 42, 1413–1422.

[5] Grounds, O.; Zeller, M.; Rosokha, S. V. Structural preferences in strong anion–π and halogen-bonded complexes: π- and σ-holes vs. frontier orbitals interaction. *New J. Chem.* **2018**, 42, 10572–10583.

[6] Jia, C.; Miao, H.; Hay, B. P. Crystal Structure Evidence for the Directionality of Lone Pair–π interactions: Fact or Fiction? *Cryst. Growth Des.* **2019**, 19, 6806–6821.

[7] Mooibroek, T. J.; Gamez, P.; Reedijk, J. Lone pair–π interactions: a new supramolecular bond? *CrystEngComm* **2008**, 10, 1501–1515.

[8] Sing, S. K.; Das, A. The n $\longrightarrow$ π^* interaction: a rapidly emerging non-covalent interaction. *Phys. Chem. Chem. Phys.* **2015**, 17, 9596–9612.

[9] Newberry, R. W.; Raines, R. T. The n$\longrightarrow\pi^*$ Interaction. *Acc. Chem. Res.* **2017**, 50, 1838–1846.

[10] Schottel, B. L.; Chifotides, H. T. Dunbar, K. R. Anion-π interactions. *Chem. Soc. Rev.* **2008**, 37, 68–83.

[11] Gamez, P.; Mooibroek, T. J.; Teat, S. J.; Reedijk, J. Anion Binding Involving π-Acidic Heteroaromatic Rings. *Acc. Chem. Res.* **2007**, 40, 435–444.

[12] Wang, D.-X.; Wang, M.-X. Anion–π Interactions: Generality, Binding Strength, and Structure. *J. Am. Chem. Soc.* **2013**, 135, 892–897.

[13] Lucas, X.; Bauzá, A.; Frontera, A.; Quiñonero, D. A thorough anion–π interaction study in biomolecules: on the importance of cooperativity effects. *Chem. Sci.* **2016**, 7, 1038–1050.

[14] Savastano, M.; Bazzicalupi, C.; García, C.; Gellini, C.; Lopez de la Torre, M. D.; Mariani, P.; Pinchierri, F.; Bianchi, A.; Melguizo, M. Iodide and triiodide anion complexes involving anion-π interactions with a tetrazine-based receptor. *Dalton Trans.* **2017**, 46, 4518–4529.

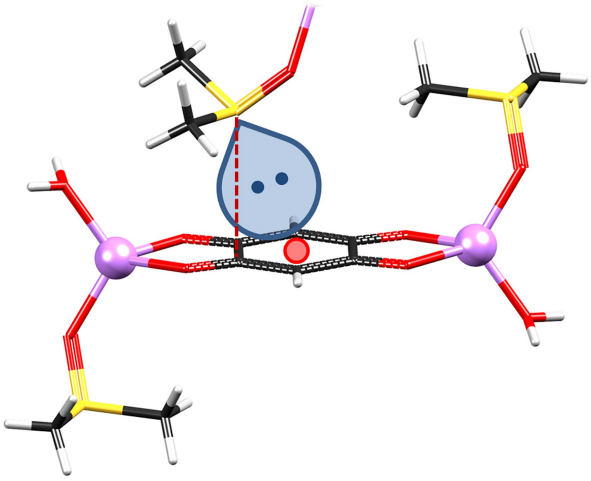
[15] Molčanov, K.; Mali, G.; Grdadolnik, J.; Stare, J.; Stilinović, V.; Kojić-Prodić, B. Iodide···π interactions of perhalogenated quinoid rings in co-

crystals with organic bases. *Cryst. Growth Des.* **2018**, 18, 5182–5193.
[16] Kepler, S.; Zeller, M.; Rosokha, S. V. Anion–π Complexes of Halides with p-Benzoquinones: Structures, Thermodynamics, and Criteria of Charge Transfer to Electron Transfer Transition. *J. Am. Chem. Soc.* **2019**, 141, 9338–9348.
[17] Wilson, J.; Maxson, T.; Wright, I.; Zeller, M.; Rosokha, S. V. Diversity and uniformity in anion–π complexes of thiocyanate with aromatic, olefinic and quinoidal π-acceptors. *Dalton Trans.* **2020**, 49, 8734–8743, DOI: 10.1039/d0dt01654c.
[18] Vuković, V.; Molčanov, K.; Jelsch, C.; Wenger, E.; Krawczuk, A.; Jurić, M.; Androš Dubraja, L.; Kojić-Prodić, B. Malleable Electronic Structure of Chloranilic Acid and Its Species Determined by X-ray Charge Density Studies. *Cryst. Growth Des.* **2019**, 19, 2802–2810.
[19] Molčanov, K.; Jelsch, C.; Landeros, B.; Hernandez-Trujillo, J.; Wenger, E.; Stilinović, V.; Kojić-Prodić, B.; Escudero-Adan, E. Partially covalent two-electron/multicentric bonding between semiquinone radicals. *Cryst. Growth Des.* **2019**, 19, 391–402.
[20] Molčanov, K.; Milašinović, V.; Kojić-Prodić, B. Contribution of different crystal packing forces in π-stacking: from non-covalent to covalent multicentric bonding. *Cryst. Growth Des.* **2019**, 19, 5967–5980.
[21] Molčanov, K.; Kojić-Prodić, B. Towards understanding of π-stacking interactions of non-aromatic rings. *IUCrJ* **2019**, 6, 156–165.
[22] Molčanov, K.; Kojić-Prodić, B.; Meden, A. π-Stacking of quinoid rings in crystals of alkali diaqua hydrogen chloranilates. *CrystEngComm* **2009**, 11, 1407–1415.
[23] Molčanov, K.; Kojić-Prodić, B.; Babić, D.; Stare, J. Face-to-face stacking of dianionic quinoid rings in crystals of alkali salts of 2,5-dihydroxyquinone in view of π –system polarization. *CrystEngComm* **2013**, 15, 135–143.
[24] Molčanov, K.; Kojić-Prodić, B.; Meden, A. Unique electronic and structural properties of 1,4-benzoquinones: crystallochemistry of alkali chloranilate hydrates. *Croat. Chem. Acta* **2009**, 82, 387–396.
[25] Molčanov, K.; Kojić-Prodić, B. Salts and cocrystals of chloranilic acid with organic bases: is it possible to predict a salt formation? *CrystEngComm* **2010**, 12, 925–939.
[26] Molčanov, K.; Sabljić, I.; Kojić-Prodić, B. Face-to-face π-stacking in the multicomponent crystals of chloranilic acid, alkali hydrogenchloranilates, and water. *CrystEngComm* **2011**, 13, 4211–4217.
[27] Molčanov, K.; Kojić-Prodić, B. Face-to-face stacking of quinoid rings in crystals of alkali salts of bromanilic acid. *Acta Crystallogr.* **2012**, B68, 57–65.
[28] Molčanov, K.; Jelsch, C.; Wenger, E.; Stare, J.; Madsen, A. Ø.; Kojić-Prodić, B. Experimental evidence of 3-centre, 2-electron covalent bond character of the central O-H-O fragment on the Zundel cation in crystals of Zundel nitranilate tetrahydrate. *CrystEngComm* **2017**, 19, 3898–3901.
[29] Milašinović, V.; Molčanov, K. Nitranilic acid as a basis for construction of coordination polymers: from discrete monomers to 3D networks. *CrystEngComm* **2019**, 21, 2962–2969.
[30] Milašinović, V.; Molčanov, K. Alkali salts of nitranilic and cyanochloranilic acids. *Croat. Chem. Acta* **2018**, 91, 177–186.
[31] Molčanov, K.; Stare, J.; Kojić-Prodić, B.; Lecomte, C.; Dahaoui, S.; Jelsch, C.; Wenger, E.; Šantić A.; Zarychta, B. A polar/π model of interactions explains face-to-face stacked quinoid rings: a case study of the crystal of potassium hydrogen chloranilate dehydrate. *CrystEngComm* **2015**, 17, 8645–8656.
[32] Vuković, V.; Leduc, T.; Jelić-Matošević, Z.; Didierjean, C.; Favier, F.; Guillot, B.; Jelsch, C. A rush to explore protein-ligand electrostatic interaction energy with *Charger*. *Acta Cryst*, **2021**, D77, 1292–1304.
[33] Shahbazian, S. Why Bond Critical Points Are Not "Bond" Critical Points. *Chem. Eur. J.* **2018**, 24, 5401–5405.
[34] Bader, R. F. W. Bond Paths Are Not Chemical Bonds. *J. Phys. Chem.* **2009**, A113, 10391–10396.
[35] Wick, C. R.; Clark, T. On bond-critical points in QTAIM and weak interactions. *J. Mol. Model.* **2018**, 24, 142.
[36] Xiang, J.; Chang, C.: Li, M.; Yuan, L.; Sun, J. A Novel Coordination Polymer as Positive Electrode Material for Lithium Ion Battery. *Cryst. Growth Des.* **2008**, 8, 280–282.

[37] Bruker. SAINT. Bruker AXS Inc., Madison, Wisconsin, USA, 2012.
[38] Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. OLEX2: A Complete Structure Solution, Refinement and Analysis Program. *J. Appl. Crystallogr.* **2009**, 42, 339–341.
[39] Sheldrick, G. M. A short history of *SHELX*. *Acta Cryst.* **2008**, A64, 112–122.
[40] Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Cryst.* **2015**, C71, 3–8.
[41] Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *Mercury CSD* 2.0 – new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008**, 41, 466–470.
[42] Persistence of Vision Pty. Ltd. Persistence of VisionTM Raytracer. Persistence of Vision Pty. Ltd., Williamstown, Australia, 2004. http://www.povray.org/ (last access 2020-04-17).
[43] Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. Advances in protein and small-molecule charge-density refinement methods using MoPro. *J. Appl. Crystallogr.* **2005**, 38, 38–54.
[44] Blessing, R. H. An empirical correction for absorption anisotropy. *Acta Crystallogr.* **1995**, A51, 33–38.
[45] Madsen, A. Ø. SHADE web server for estimation of hydrogen anisotropic displacement parameters. *J. Appl. Crystallogr.* **2006**, 39, 757–758.
[46] Madsen, A. Ø.; Hoser, A. A. A simple approach to estimate isotropic displacement parameters for hydrogen atoms. *Acta Crystallogr.* **2015**, A71, 169–174.
[47] Allen, F. H.; Bruno, I. J. Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Crystallogr.* **2010**, B66, 380–386.
[48] Domagała, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An improved experimental databank of transferable multipolar atom models - ELMAM2. Construction details and applications. *Acta Crystallogr.* **2012**, A68, 337–351.
[49] Abrahams, S. C.; Keve, E. T. Normal probability plot analysis of error in measured and derived quantities and standard deviations. *Acta Crystallogr.* **1971**, A27, 157–165.
[50] Zhurov, V. V.; Zhurova, E. A.; Pinkerton, A. A. Optimization and evaluation of data quality for charge density studies. *J. Appl. Crystallogr.* **2008**, 41, 340–349.
[51] Meindl, K.; Henn, J. Foundations of residual-density analysis. *Acta Crystallogr.* **2008**, A64, 404–418.
[52] Meindl., K. A. C. *Residual density validation and the structure of Labyrinthopeptin A2*, PhD Thesis. Mathematisch-Naturwissenschaftliche Fakultät, Georg-August-Universität Göttingen, October 2008.
[53] Tang, W.; Sanville E.; Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *J. Phys. Condens. Matter* **2009**, 21, 084204.
[54] Howard, S. T.; Lamarche, O. Description of covalent bond orders using the charge density topology. *J. Phys. Org. Chem.* **2003**, 16, 133–141.
[55] Tsirelson, V. G.; Bartashevich, E. V.; Stash, A. I.; Potemkin, V. A. Determination of covalent bond orders and atomic valence indices using topological features of the experimental electron density. *Acta Crystallogr.* **2007**, B63, 142–150.
[56] Bartashevich, E.; Nikulov, D.; Vener, M.; Tsirelson, V. QTAIMC study of the X–H/H···O bond order indices (X=O, N, C) in molecular systems. *Comput. Theor. Chem.* **2011**, 973, 33–39.
[57] Guillot, B.; Enrique, E.; Huder, L.; Jelsch, C. MoProViewer: a tool to study proteins from a charge density science perspective. *Acta Crystallogr.* **2014**, A70, C279.
[58] Fournier, B.; Guillot, B.; Lecomte, C.; Escudero-Adán, E. C.; Jelsch, C. A Method to Estimate Statistical Errors of Properties Derived from Charge-Density Modelling. *Acta Crystallogr.* **2018**, A74, 170–183.
[59] Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C.; Jupyter Development Team, „Jupyter Notebooks – a publishing format for reproducible computational workflows". *Proceedings of the 20th International Conference on Electronic Publishing*. F. Loizides & B. Schmidt, eds. 2016.
[60] Python Software Foundation, Python, 2001, `https://www.python.org/`. (Last access 2020-04-17).

[61] McKinney, W. „Data Structures for Statistical Computing in Python". *Proceedings of the 9th Python in Science Conference.* S. van der Walt & J. Millman, eds. 2010, pp. 51–56.
[62] Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco P.; Noël, Y.; Causà, M.; Rérat, M; Kirtman, B. CRYSTAL14: A program for the ab initio investigation of crystalline solids. *Int. J. Quantum Chem.* **2014**, 114, 1287–1317.
[63] Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I. J.; D'Arco, P.; Llunell, M.; Causà, M.; Noël, Y. *CRYSTAL14 User's Manual.* Università di Torino, Torino, Italy, 2014.
[64] Gatti, C.; Saunders, V. R.; Roetti, C. Crystal field effects on the topological properties of the electron density in molecular crystals: The case of urea. *J. Chem. Phys.* **1994**, 101, 10686.
[65] Heyd, J.; Peralta, J. E.; Scuseria, G. E.; Martin, R. L. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Chem. Phys.* **2005**, 123, 174101.
[66] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. von R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *J. Comput. Chem.* **1983**, 4, 294–301.
[67] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2009.
[68] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, 98, 5648–5652.
[69] Papajak, E.; Zheng, J.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of DiffuseBasis Functions. *J. Chem. Theory and Comput.* **2011**, **7,** 3027–3034.
[70] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, **132**, 154104.
[71] Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.*, **1970**, 19, 553.
[72] Yanai, T.; Tew, D.; Handy, N. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, 393, 51–57.

GRAPHICAL ABSTRACT



An unusual intermolecular interaction between a lone pair and an electron-rich π-electron system of a quinoid dianion was studied in detail by a combination of X-ray charge density and quantum chemical modelling. The presence of a bonding critical point is largely a result of a generally attractive electrostatic interaction the quinoid dianion and DMSO, and not so much of a bonding interaction between close atoms.

C Supplementary Material for Cyclopentadienyl‴NiP_3

Table C.1. Details on the gathered crystallographic data and charge density refinement of investigated compounds.

Compound	Cp‴NiP_3
Empirical formula	$C_{17}H_{29}P_3Ni$
Empirical mass / (g/mol)	356.812
Crystal dimensions / mm^3	$0.29 \times 0.27 \times 0.23$
Space group	$P2_1/c$
a / Å	9.2644(1)
b / Å	14.5889(1)
c / Å	14.5261(1)
α / °	90
β / °	105.935(1)
γ / °	90
V / Å	1887.87(3)
D_{calc} / (g/cm^3)	1.255
μ / mm^{-1}	1.270
Θ range / °	2.68 - 41.15
T / K	100.00(10)
λ / Å	0.71073
Diffractometer type	Rigaku SuperNova
h, k, l range	$17 \geq h \geq -17$
	$26 \geq k \geq -26$
	$26 \geq l \geq -26$
Resolution / Å	0.54
Total reflections	279297
Independent reflections	12441
Reflections with $I > 2\sigma$	11629
Absorption correction	Empirical
T_{min}	0.435

The table continues on the following page.

The continuation of table C.1 from the previous page.

Compound	Cp'''NiP_3
T_{max}	1.000
Statistical weighting scheme	$[2\sigma(I_{obs})]^{-}1$
Completeness (%)	99.94
R_{int} (%)	2.93
$R(F)$ (%)	1.548
$R(F^2)$ (%)	1.427
Goodness-of-fit	1.062
Number of parameters	807
Number of constraints/restraints	127
$\Delta\rho_{min}$ / e Å^{-3}	-0.28211
$\Delta\rho_{max}$ / e Å^{-3}	0.29244

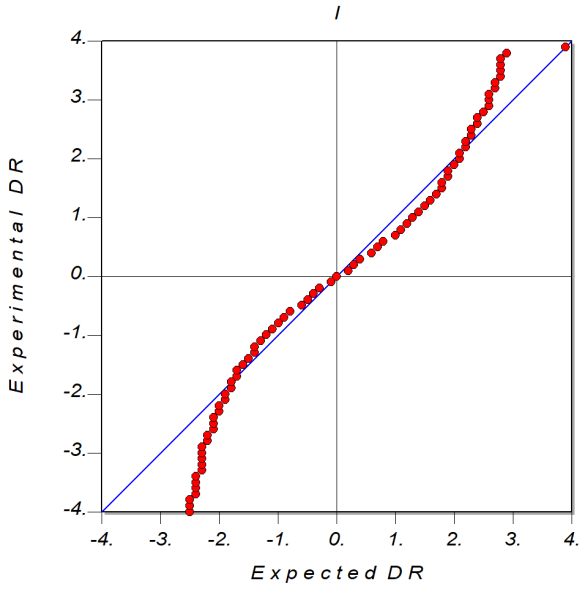


Figure C.1. The observed versus theoretical quantile (q-q) plot for the Cp'''NiP_3 structure. The experimental *DR* values represent..., while the theoretical *DR* values are their equivalents from a standard normal distribution. An ideal dataset would have all the points align on the $y = x$ line.

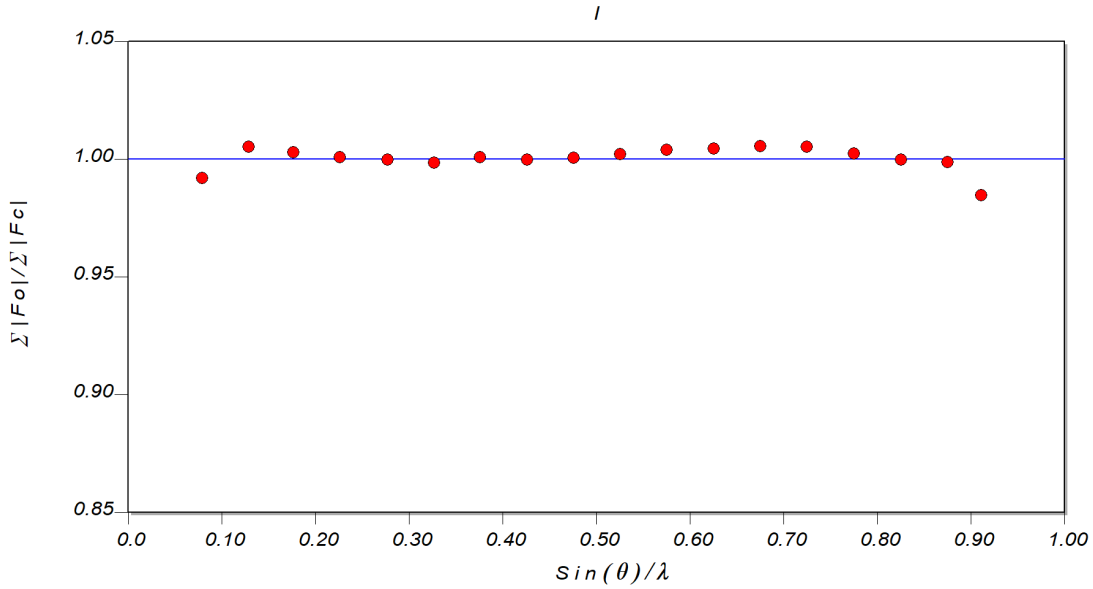


Figure C.2. The ratio of observed to calculated structure factors $(I_{\text{obs}})/I_{\text{calc}})$ as a function of reciprocal resolution. An ideal dataset would have this ratio equal to 1 (the blue line).

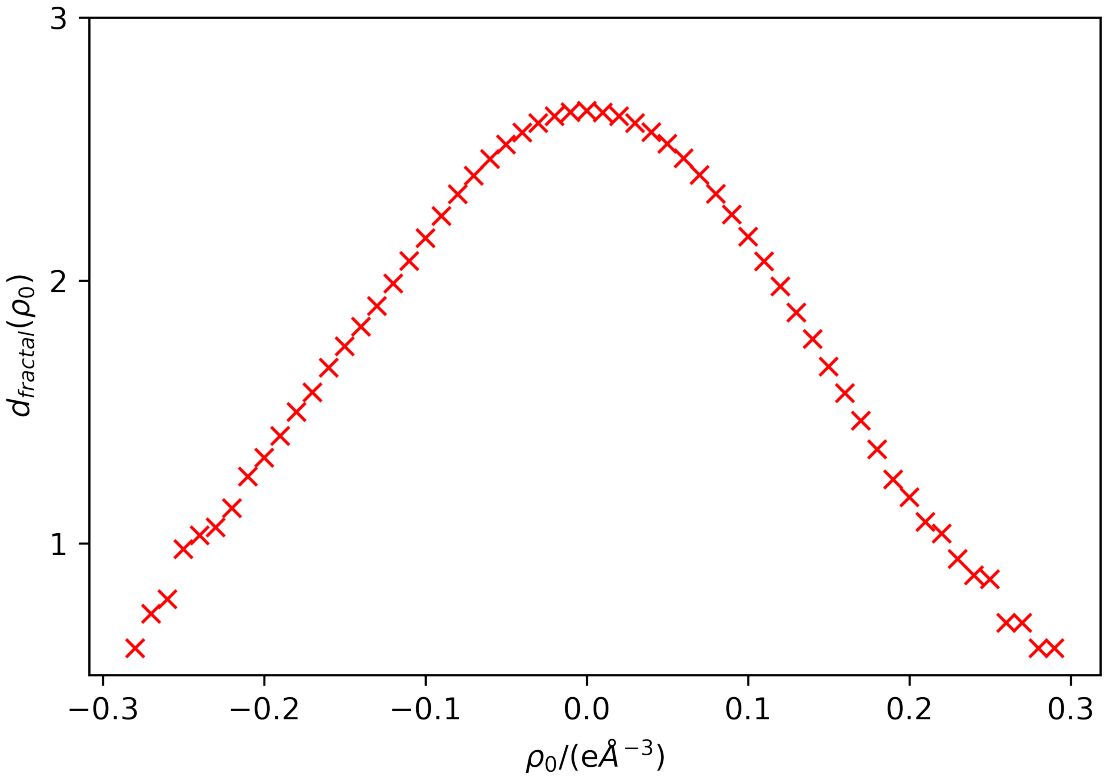
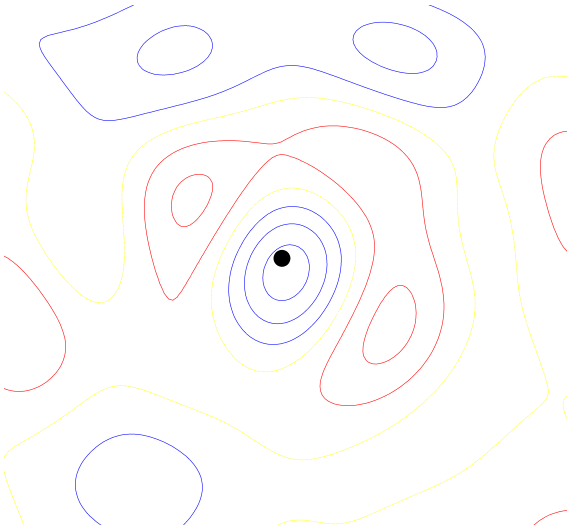
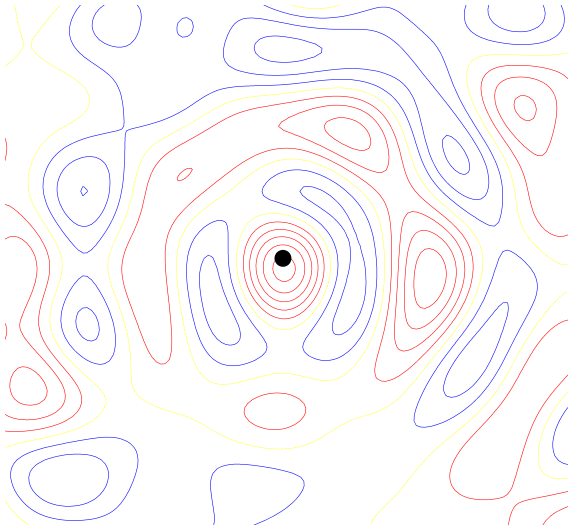


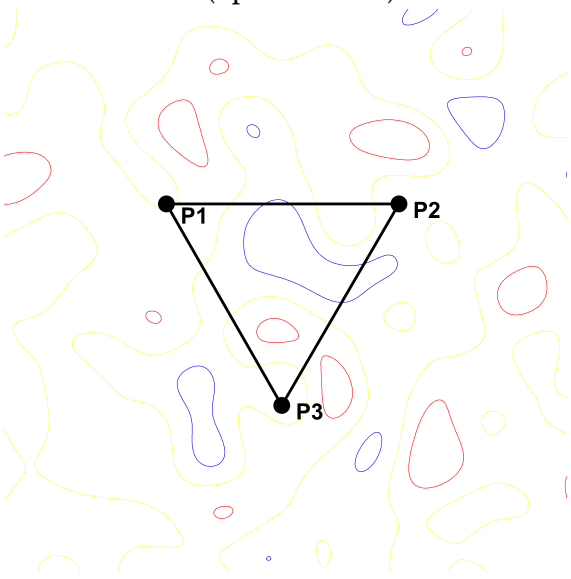
Figure C.3. The fractal dimension d_{fractal} as a function of the residual density isosurface value ρ_0 for the Cp‴NiP_3 residual density. A residual density coming from purely Gaussian noise would result in a parabola-shaped plot with a maximum at $(0,\ 3)$.



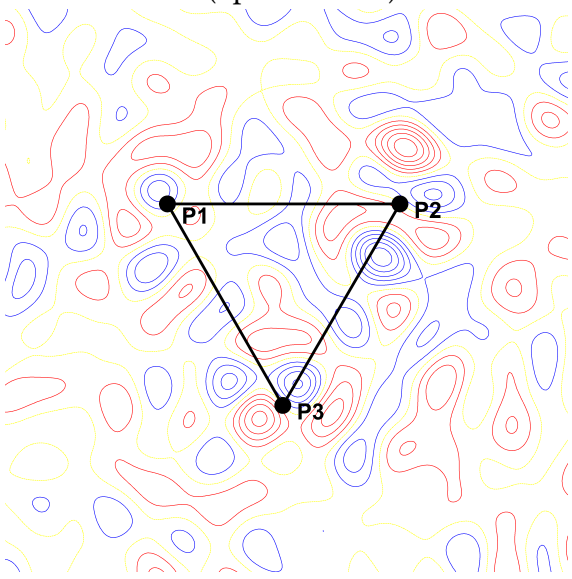
Ni (up to 0.7 Å^{-1}).



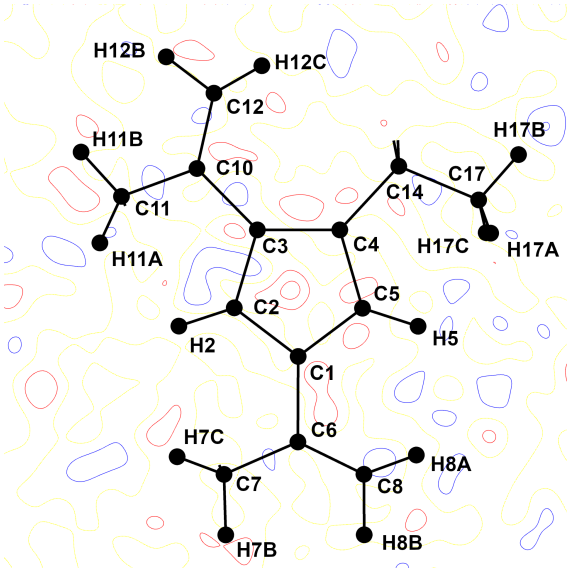
Ni (up to 1.0 Å^{-1}).



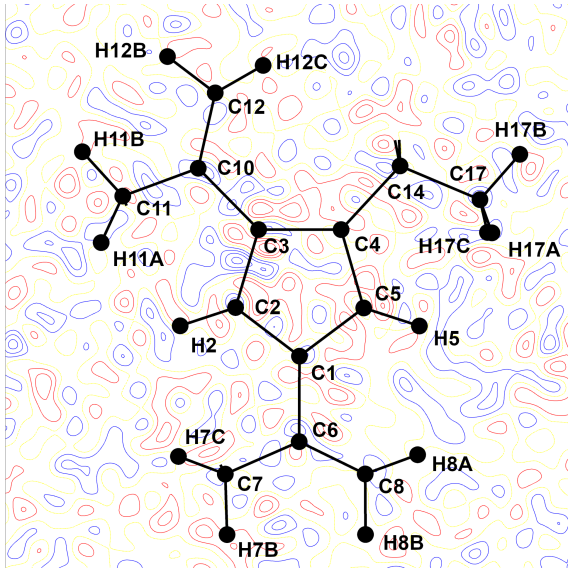
Cyclo-P$_3$. (up to 0.7 Å^{-1}).



Cyclo-P$_3$. (up to 1.0 Å^{-1}).



Cp‴ (up to 0.7 Å^{-1}).



Cp‴ (up to 1.0 Å^{-1}).

Figure C.4. The two-dimensional residual density plots of various components of Cp‴NiP$_3$.

sΘ⊕

CALCULS ÉLECTROSTATIQUES RAPIDES DANS L'ANALYSE ÉNERGÉTIQUE

Développement de la méthode, applications aux études de cristaux organiques et organométalliques et de complexes protéine/ligand

Mots clés :

Énergie électrostatique, cristallographie quantique, glutathion transférase, quinones, complexes organométalliques.

Résumé :

La nature électrostatique des atomes est généralement prise en compte pour explorer les interactions intermoléculaires. Le modèle multipolaire de Hansen-Coppens est largement utilisé pour obtenir une description détaillée de la distribution des électrons. Il s'agit donc d'un point de départ pour étudier le comportement de systèmes chimiques. Le choix du système donne accès à des informations microscopiques qui permettent d'élucider les propriétés des matériaux à l'échelle macroscopique.

L'une de ces propriétés est l'énergie d'interaction électrostatique, qui indique la force avec laquelle deux charges s'attirent ou se repoussent. Le modèle multipolaire donne une image claire de la distribution des charges à l'intérieur des molécules. Les énergies obtenues à partir de ce modèle sont plus fiables que celles provenant de méthodes traditionnelles (comme les charges ponctuelles). Des méthodes existent déjà pour calculer l'énergie d'interaction électrostatique à partir du modèle multipolaire, mais leur point faible est qu'elles prennent du temps pour calculer les intégrales sous-jacentes.

La méthode discutée dans cette thèse (potentiel analytique exact / moments pseudo-multipolaires, aEP/pMM) propose une intégration analytique rapide à la place de l'intégration numérique de l'ancienne méthode. Cette nouvelle méthode a été intégrée dans le programme `Charger`. Le gain de vitesse provient du remplacement d'une cubature 3D par une intégrale unidimensionnelle, grâce aux propriétés mathématiques du modèle multipolaire. Ce gain varie d'un facteur 4 à 200, et les énergies résultantes sont presque indiscernables de celles obtenues par l'ancienne méthode.

L'objectif était également d'estimer le potentiel de `Charger` en termes d'application. Il a été testé sur des complexes protéine-ligand, des molécules organiques et des complexes métal-organiques impliquant de nombreux éléments des trois premières rangées du tableau périodique. Un complexe organométallique contenant des atomes de nickel dans son cœur a également été analysé. Dans toutes ces applications, `Charger` domine sur l'ancienne approche en termes de temps de calcul, tout en conservant une précision exceptionnelle.

L'analyse du complexe protéine/ligand a permis de distinguer les environnements favorables et défavorables à la fixation du ligand. Elle a également révélé un candidat possible à la mutation qui confère un pouvoir prédictif à l'approche `Charger`. Le complexe métal-organique contenait un contact particulier que `Charger` a aidé dans sa caractérisation. Il a également permis le calcul de l'énergie électrostatique de la maille cristalline pour un composé organique et un composé organométallique. L'auteur espère que la nouvelle approche aEP/pMM recevra un accueil positif et contribuera à de nombreuses découvertes dans divers domaines qui reposent sur la modélisation moléculaire.

Fast electrostatic calculations in energy analysis

Method development, applications in investigating organic and organometalic crystals, and to protein/ligand complexes

Keywords:

Electrostatic energy, quantum crystallography, glutathione transferase, quinones, organometallic complexes.

Abstract:

One usually considers the electrostatic nature of atoms to explore intermolecular interactions. The Hansen-Coppens multipolar model is widely used to bring a detailed view of electron distribution on an atomic scale. It is thus a common starting point to investigate chemical behaviour of selected systems, giving access to microscopic insights that elucidate material properties on the large scale.

One such property is the electrostatic interaction energy, which tells how strongly two charges attract or repel each other. The multipolar model brings forward a clear picture of charge distributions inside molecules, so a calculation of electrostatic energy based on it provides a better estimate than by using traditional methods (simplifications like point charges). The methods already exist to calculate electrostatic interaction energy, but their weak point is that they take their time to calculate the underlying integrals.

The method discussed in this thesis (analytical exact potential / pseudo-multipolar moments, aEP/pMM) replaces the slow numerical integration inherent to the older method with analytical integration. It was integrated in the program `Charger`. The boost of speed comes from replacing a 3D cubature with a one-dimensional integral, due to favourable mathematical properties of the underlying multipolar model. The improvement in speed goes from 4 to 200 times, and the resulting energies are almost indistinguishable from those obtained by the older method.

The goal was also to see how wide one can cast the net of this new `Charger` approach. It was tested on protein/ligand complexes, organic molecules and metal-organic complexes involving many elements of the first three rows in the periodic table. It was also applied on an organometallic complex containing nickel atoms in its core. In all of these applications, `Charger` reigned supreme over the old approach in terms of computational time, while guarding exceptional accuracy.

The application on the protein/ligand complex pointed to residues which are the major contributors and detractors to ligand binding. It also revealed a possible mutation candidate that could show the predictive power of the `Charger` approach. The metal-organic complex contained a peculiar contact: `Charger` helped to characterise it. `Charger` also unveiled the electrostatic energy of crystal packing for an organic and an organometallic compound. The author hopes that the new aEP/pMM approach will encounter a positive reception and contribute to many an interesting finding across various fields that rely on molecular modeling.

Résumé :

La nature électrostatique des atomes est généralement prise en compte pour explorer les interactions intermoléculaires. Le modèle multipolaire de Hansen-Coppens est largement utilisé pour obtenir une description détaillée de la distribution des électrons. Il s'agit donc d'un point de départ pour étudier le comportement de systèmes chimiques. Le choix du système donne accès à des informations microscopiques qui permettent d'élucider les propriétés des matériaux à l'échelle macroscopique.

L'une de ces propriétés est l'énergie d'interaction électrostatique, qui indique la force avec laquelle deux charges s'attirent ou se repoussent. Le modèle multipolaire donne une image claire de la distribution des charges à l'intérieur des molécules. Les énergies obtenues à partir de ce modèle sont plus fiables que celles provenant de méthodes traditionnelles (comme les charges ponctuelles). Des méthodes existent déjà pour calculer l'énergie d'interaction électrostatique à partir du modèle multipolaire, mais leur point faible est qu'elles prennent du temps pour calculer les intégrales sous-jacentes.

La méthode discutée dans cette thèse (potentiel analytique exact / moments pseudo-multipolaires, aEP/pMM) propose une intégration analytique rapide à la place de l'intégration numérique de l'ancienne méthode. Cette nouvelle méthode a été intégrée dans le programme `Charger`. Le gain de vitesse provient du remplacement d'une cubature 3D par une intégrale unidimensionnelle, grâce aux propriétés mathématiques du modèle multipolaire. Ce gain varie d'un facteur 4 à 200, et les énergies résultantes sont presque indiscernables de celles obtenues par l'ancienne méthode.

L'objectif était également d'estimer le potentiel de `Charger` en termes d'application. Il a été testé sur des complexes protéine-ligand, des molécules organiques et des complexes métal-organiques impliquant de nombreux éléments des trois premières rangées du tableau périodique. Un complexe organométallique contenant des atomes de nickel dans son cœur a également été analysé. Dans toutes ces applications, `Charger` domine sur l'ancienne approche en termes de temps de calcul, tout en conservant une précision exceptionnelle.

L'analyse du complexe protéine/ligand a permis de distinguer les environnements favorables et défavorables à la fixation du ligand. Elle a également révélé un candidat possible à la mutation qui confère un pouvoir prédictif à l'approche `Charger`. Le complexe métal-organique contenait un contact particulier que `Charger` a aidé dans sa caractérisation. Il a également permis le calcul de l'énergie électrostatique de la maille cristalline pour un composé organique et un composé organométallique. L'auteur espère que la nouvelle approche aEP/pMM recevra un accueil positif et contribuera à de nombreuses découvertes dans divers domaines qui reposent sur la modélisation moléculaire.

Abstract:

One usually considers the electrostatic nature of atoms to explore intermolecular interactions. The Hansen-Coppens multipolar model is widely used to bring a detailed view of electron distribution on an atomic scale. It is thus a common starting point to investigate chemical behaviour of selected systems, giving access to microscopic insights that elucidate material properties on the large scale.

One such property is the electrostatic interaction energy, which tells how strongly two charges attract or repel each other. The multipolar model brings forward a clear picture of charge distributions inside molecules, so a calculation of electrostatic energy based on it provides a better estimate than by using traditional methods (simplifications like point charges). The methods already exist to calculate electrostatic interaction energy, but their weak point is that they take their time to calculate the underlying integrals.

The method discussed in this thesis (analytical exact potential / pseudo-multipolar moments, aEP/pMM) replaces the slow numerical integration inherent to the older method with analytical integration. It was integrated in the program `Charger`. The boost of speed comes from replacing a 3D cubature with a one-dimensional integral, due to favourable mathematical properties of the underlying multipolar model. The improvement in speed goes from 4 to 200 times, and the resulting energies are almost indistinguishable from those obtained by the older method.

The goal was also to see how wide one can cast the net of this new `Charger` approach. It was tested on protein/ligand complexes, organic molecules and metal-organic complexes involving many elements of the first three rows in the periodic table. It was also applied on an organometallic complex containing nickel atoms in its core. In all of these applications, `Charger` reigned supreme over the old approach in terms of computational time, while guarding exceptional accuracy.

The application on the protein/ligand complex pointed to residues which are the major contributors and detractors to ligand binding. It also revealed a possible mutation candidate that could show the predictive power of the `Charger` approach. The metal-organic complex contained a peculiar contact: `Charger` helped to characterise it. `Charger` also unveiled the electrostatic energy of crystal packing for an organic and an organometallic compound. The author hopes that the new aEP/pMM approach will encounter a positive reception and contribute to many an interesting finding across various fields that rely on molecular modeling.